Dissertation supervisors:

Prof. Dr. ir. Peter Goethals
Research Group Aquatic Ecology (AECO)
Department of Animal Sciences and Aquatic Ecology
Faculty of Bioscience Engineering, Ghent University, Belgium

Prof. Dr. ir. Ingmar Nopens
BIOMATH Research Unit
Department of Data Analysis and Mathematical Modelling
Faculty of Bioscience Engineering, Ghent University, Belgium

Members of the examination committee

Prof. Dr. ir Filip Tack
Department of Green Chemistry and Technology
Faculty of Bioscience Engineering, Ghent University, Belgium

Prof. Dr. ir. Pascal Boeckx
Department of Green Chemistry and Technology
Faculty of Bioscience Engineering, Ghent University, Belgium

Dr. Tuan Long Ho
Research Group Aquatic Ecology (AECO)
Department of Animal Sciences and Aquatic Ecology
Faculty of Bioscience Engineering, Ghent University, Belgium

Prof. Dr. Olivier Thas
Department of Mathematics and Statistics
Faculty of Sciences, Hasselt University, Belgium

Prof. Dr. Eng. Felipe Cisneros
PROMAS, Programa para el manejo del agua y del suelo
Universidad de Cuenca, Ecuador

Prof. Dr. Luis Dominguez
CADS, Centro del Agua y Desarrollo Sustentable
Escuela Superior Politécnica del Litoral (ESPOL), Ecuador

Dean of the faculty of Bioscience Engineering, Ghent University

Prof. Dr. ir. Marc Van Meirvenne

Rector of Ghent University

Prof. Dr. ir. Rik Van de Walle

Faculty of Bioscience Engineering

Rubén Fernando Jerves Cobo

Integrated water system modelling to support water management in the Cuenca Basin

Thesis submitted in fulfillment of the requirements for the degree of
Doctor (Ph.D) in Applied Biological Sciences
Academic year 2018-2019

Dutch translation of the title:

Geïntegreerde watersysteemmodellering ter ondersteuning van het waterbeheer in het Cuenca-stroombekken

Please refer to this work as follows:

Jerves-Cobo, Rubén. 2019. Integrated water system modelling to support water management in the Cuenca Basin. PhD thesis. Ghent University, Ghent, Belgium.

ISBN – number: **9789463571975**

This PhD research was supported by the VLIR-UOS IUC Programme - University of Cuenca and the VLIR-UOS Ecuador Biodiversity Network Project.

Acknowledgements

First, I would like to express my sincere gratitude to my promoters Prof. Peter Goethals and Prof. Ingmar Nopens for giving me the opportunity to undertake a PhD with their support and guidance. Thank you for your teaching, feedback, discussions and valuable insights that enriched and improved this research, and culminated in the finishing of this PhD thesis.

My whole-hearted appreciation to Prof. Felipe Cisneros, head of the research unit at the Water and Soil Management Program (PROMAS) of the University of Cuenca, for advising, encouraging and trusting me on this journey. Similarly, my gratitude goes to Prof. Fabian Leon, Dean of the Chemical Faculty of the University of Cuenca and Ecuadorian Coordinator of the VLIR-UOS-Programme, for believing in and supporting me during this process.

I extended my truly appreciation to my examination committee, Prof. ir Filip Tack (Ghent University), Prof. ir. Pascal Boeckx (Ghent University), Dr. Tuan Long Ho (Ghent University), Prof. Dr. Olivier Thas (Hasselt University), Prof. Felipe Cisneros (University of Cuenca) and Prof. Luis Dominguez (ESPOL). Thank you for your carefully reading, valuables comments and corrections of this dissertation. I would like to send special thanks to Prof. Filip Tack for being chairman of my PhD defense.

I am very grateful to the Inter University Cooperation Program from the Flemish Interuniversity Council (VLIR-UOS), with special reference to the IUC - University of Cuenca and the Ecuador Biodiversity Network projects, which financed my stays in Ghent and provided the means for the sampling campaigns.

I render my thanks to the Water Supply and Wastewater Management Municipal Company ETAPA-EP and Electro Generator of the Austro Company (ELECAUSTRO) for permitting me to take samples on their grounds and for supplying me with information regarding the flow of the Cuenca Rivers system. Additionally, ETAPA-EP shared valuable information with me regarding the wastewater treatment plants. Likewise, my appreciation goes to the Ecuadorian Environmental Ministry for facilitating the sampling permits.

Recognition is due to the University of Cuenca, the Salesian Polytechnic University (UPS), the University of Azuay (UDA) and the Technical University of the North (UTN) for their support during my research. My special thanks to Prof. Fabian Carrasco (Former-Rector of the U Cuenca), Prof. Pablo Vanegas (Rector of the U Cuenca), Prof. SDB. Javier Herran (Rector of the UPS), Prof. Cesar Vazquez (Vice-rector of the UPS) and Prof. Gustavo Chacon (Former-Research Dean of the UDA).

I extend my appreciation to my former and current colleagues from the Laboratory of Environmental Toxicology and Aquatic Ecology: Sigrid, Marianne, Nancy, Euri, Tiptiwa, Wout, Niels, Stijn, Long, Tu Tri, Shewit, Daniel, Lenin, Gisèle, Jolien, Marc, Sylvie, Veerle, Gert, Tien, Ratha, Sacha, Selamawit, Natalia and Jana, for your support, your assistance and for sharing your time experiences during my PhD. My special thanks to Dr. Koen Lock for the identification of invertebrates and to Jana Van Butsel for her support during the sampling campaigns.

I offer similar gratitude to my colleagues in the Department of Data Analysis and Mathematical Modelling (BIOMATH): Chaim, Jan and Jenna who also supported my research. More special thanks go to Dr. Youri Amerlinck and Dr. Lorenzo Benedetti for sharing their expertise during the last step of my research.

I extend my appreciation of the support given to me during the sampling campaigns to the staff from PROMAS and the Sanitary Laboratory of the University of Cuenca, who helped in field and laboratory work. Additional special thanks go to Guillermina, Oscar and Gabriela.

I would also say a special thank you to my Ecuadorian and international fellows Jessica, Paulina, Gonzalo, Karol, Maria Helen, Margarita, Gonzalo and Fien for the good moments together, for the moral support and for making my stays in Ghent manageable. Similar appreciation goes to my landlord Machteld, who provided a space for my stays in Ghent. I also thank Annick and Madina from the Department of Research Affairs of Ghent University, who were the responsible for the monthly disbursements of my scholarship and the arrangements of my travels. A special acknowledgment to my friend and former teacher, Helena from the US, who helped me to improve my English.

This work is dedicated to my wife Jimena and my sons Fernando, Rubén Andres and Mateo, whom I would also like to give my eternal gratitude for encouraging me to reach my goals in life, and especially during these last four years. Jimena, thank you for your understanding, generosity, patience and love, and for taking care of our family while I was abroad, far from our home and our family. I also want to say thank you to my parents Ruben and Maria Elena, who with their example, have taught me honesty, responsibility, the courage to overcome difficulties and to fight for my goals. Similarly, appreciation goes to my siblings Caty and Alex, and my nieces Carolina and Pamela for boosting my confidence since I first embarked upon this project.

Finally, from the depth of my soul and heart, my endless thankfulness to God for everything that I have received in my life.

Rubén Jerves Cobo

Ghent, 24 April 2019

Table of contents

List of symbols and notations

Symbol	Description
α	Intercept coefficient of a GLM
β_i	Regression coefficient of a GLM
η_i	Linear predictor
μ_i	Mean of the response or dependent variable
σ	Standard deviation
ϑ^{-1}	Shape parameter of a Gamma or an Inverse Gaussian GLM
X_i	Explanatory variable of a GLM
BOD_5	Five-day biochemical oxygen demand
Cl	Chloride
COD	Chemical oxygen demand
color	True color
DO	Dissolved Oxygen
FC	Fecal coliforms
κ	Cohen's Kappa statistic
Kla_{base}	Re-aeration coefficient
$K_{gro,N2}$	Maximum specific growth rate of 2^{nd} stage nitrifiers
NaCl	Sodium chloride
NH_4	Ammonium
NO_2	Nitrite
NO_3	Nitrate
OS	Oxygen saturation
π_i	The probability that taxa *i* are present
PO_4	Phosphates
TKN	Total Kjeldahl nitrogen
S_I	Inert soluble COD
S_S	Biodegradable soluble COD
X_{ALG}	Algae and macrophytes
X_I	Particulate inert COD
X_H	Heterotrophic biomass
X_{N1}	First stage nitrifying bacteria
X_{N2}	Second stage nitrifying bacteria
X_S	Particulate organic matter
χ^2	Chi squared

List of abbreviations and acronyms

A

ABI	Andean Biotic Index
AIC	Akaike information criterion
ANNs	Artificial neural networks
ASM1	Activated Sludge Model No. 1
ASM2d	Activated Sludge Model No. 2d

B

BBNs	Bayesian belief networks
BMWP	Biological Monitoring Working Party
BMWP-Col	Biological Monitoring Working Party adapted to Colombia
BWQ	Biological Water Quality

C

CCI	Correctly classified instances
CEN	Confusion entropy of a confusion matrix
CMW	Cost matrix weights
CSTRS	Continuously stirred tank reactors in series
CSC	Cost-sensitive classifier
CSO	Combined sewer overflow
CTs	Classification trees

D

DPSIR	Drivers, pressures, state, impact and response
DTM	Decision tree model

E

EAM	Ecological assessment models
EPT	Ephemeroptera—Plecoptera—Trichoptera
ETAPA-EP	Water Supply and Wastewater Management Municipal Company of Cuenca

F

fcv	Folds cross validation
FCR	Fecal coliform regulation
FDR	False discovery rate
FN	Number of false negative

FP	Number of false positive
FWER	Family wise error rate

G

GAs	Genetic algorithms
GLM	Generalized linear models
GLUE	Generalized Likelihood Uncertainty Estimation
G-WWTP	Guangarcucho wastewater treatment plant

H

HMID	Hydro-morphological index of diversity
HSM	Habitat-suitability models
HRT	Hydraulic retention time

I

IEM	Integrated ecological model
INEC	Ecuadorian National Institute of Statistics and Census
ISO	International Organization for Standardization
IUWS model	Integrated urban wastewater system model
IRM model	Integrated river management model

K

k-fold	Number of folds (k) used in data base splitting

L

LRs	Logistic regressions

M

m a.s.l.	Meters above sea level
MeanCRS	Mean central relative sensitivity
MPN.100 mL^{-1}	Most probable number per 100 mL
MAE-Ecuador	Ministry of Environment of Ecuador

N

NST	Number of sensitive taxa

P

PCF	Pruning confidence factor

P_{FN}	Misclassification probability of classifying the samples of class i to class j subject to class j

R

RMSE	Minimum root mean square error
RTs	Regression trees
RWQM1	River water quality model one

S

Sc-1	Scenario one
Sc-2	Scenario two
Sc-3	Scenario three
Sc-4	Scenario four
SRC	Standardized regression coefficient
SWAT	Soil and Water Assessment Tool
SVMs	Support-vector machines
SWD	Shannon-Wiener index
SWO	Surface water outfalls
SENAGUA	Ecuadorian National Secretary of Water
SENPLADES	Ecuadorian National Secretariat for Planning and Development

T

TN	Number of true negative
TP	Number of true positive
TS	Tolerant score
TSol	Total Solids

U

USEPA	U.S. Environmental Protection Agency
U-WWTP	Ucubamba-wastewater treatment plant

W

Weka	Waikato Environment for Knowledge Analysis
WEST®	World Wide Engine for Simulation, Training and Automation
WQI	Water Quality Index developed by the US National Sanitation Foundation
WWTP	Wastewater treatment plant

Summary

The water quality worldwide has been degraded principally by anthropogenic causes such as intensive agriculture, industrial production, mining, untreated wastewater and urban runoff. The discharges of pollutants into the waterways not only produce changes in the water's chemical composition but also affect the morphological structure of water channels. Additionally, the waterways are affected by other morphological modifications such as erosion, changes in land use, and reservoirs. Both changes in the water composition and morphological variation in the waterways cause a degradation in the river ecology. The Cuenca River basin, located in the southern Andes of Ecuador, suffers disturbances to its aquatic ecosystem due to the presence of hydropower dams and an urbanized area that corresponds to the city of Cuenca. Several stakeholders made efforts to recover the ecological status of the Cuenca River and its tributaries. These efforts have been mainly focused on the management of the wastewater. However, the ecological status of the rivers continues to be affected by the remaining wastewater effluents and other sources of pollution. The overall aim of this dissertation is to develop an integrated river management (IRM) model for ecosystem analysis, under current and potential future conditions, that can be used as a decision-making support tool in the management of the Cuenca River basin. This study, with the help of advanced statistical methods and models, also analyzes the physicochemical, microbiological and morphological variables that could influence macroinvertebrates and provides insights in the ecological water quality of the Cuenca River system. Additionally, driving pressures such as outfalls, combined sewer overflows and discharges from wastewater treatment plants are included in a mechanistic integrated urban wastewater (IUWS) model. The results of this IUWS model are linked with ecological models to assess and to predict the water quality in the Cuenca River under current and simulated conditions.

This study was developed in the Cuenca River basin, focusing on two areas with the highest aquatic ecosystem disturbance. Thus, the urbanized area that corresponds to the city of Cuenca and its periphery, and the Machangara sub-basin that houses two dams for hydropower production were analyzed. To perform an integrated water quality assessment, data was collected during the dry and rainy seasons. Thus, information about microbiological and physicochemical conditions, as well as morphological characteristics, were collected. Similarly, aquatic macroinvertebrates were sampled and identified to the family level.

In the urban and suburban areas of the Cuenca River basin, the biological water quality was evaluated in relation to chemicals discharged through sewage outfall during dry and rainy seasons. The Andean Biotic Index (ABI) and the Biological Monitoring Working Party adapted to Colombia (BMWP-Col) were used to assess the

biological water quality. Each of these biological indexes registered higher biological water quality upstream than downstream from the city. Moreover, these indexes indicated better conditions during the rainy season, based on the presence of more sensitive macroinvertebrates families. These indexes related more to the oxygen saturation than to the five-day biological oxygen demand (BOD_5), nutrients and chloride concentrations. The relationship between the BOD_5 and nutrient concentrations with the variation of both biological indexes was clearer in the dry season. However, in some sites, these indexes were influenced more by morphological aspects than by pollutants. The ABI index was shown to be more suitable for the high Andes region.

In the Machangara River basin, the abiotic preferences of three families of the Ephemeroptera—Plecoptera—Trichoptera (EPT) taxa, *Baetidae*, *Leptoceridae and Perlidae* were evaluated. Using generalized linear models (GLMs), habitat-suitability models were constructed, analyzing the relationship between the probability of occurrence of these pollution-sensitive macroinvertebrates families and physicochemical water quality conditions. As a result, each taxon showed a different predictor for its occurrence, and with equal predictors, the range was different for each taxon. In total, eight physicochemical and morphological variables had substantial influence over the outcomes of the three models.

Similarly, the relationship between microbial pathogens and macrobenthic invertebrate taxa was examined in the Machangara River. Habitat-suitability models based on decision tree models (DTMs) were used to generate rules linking the presence and abundance of certain benthic families to microbial pathogen standards. As a result, two different and reliable models were obtained as proxy indicators in a preliminary assessment for microbial pathogen pollution in the rivers. The aforementioned DTMs provide an implied, relative, and fast method to check that two Ecuadorian water use regulations related to microbial pathogens have been fulfilled. However, this primary evaluation must accompany analyses to confirm that microbial water quality standards have been met.

Biological studies indicate substantial differences between seasons in tropical rivers. With the information collected from the rivers that pass through the city of Cuenca and its outskirts, season-specific and season-overarching models were constructed to predict biological water quality. Since the ABI is used as an indicator of biological water quality, models were developed to predict the ABI with physicochemical and morphological variables as predictors. The predictions were obtained using three kinds of GLMs: Gaussian, Gamma, and Inverse Gaussian. The season-specific models were more accurate than 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 the BOD_5, ammonium and

orthophosphate, while dissolved oxygen (DO), oxygen saturation (OS), nitrate and total solids proved to be important during the rainy season.

In order to restore the ecological water quality of the Cuenca River and its tributaries, four scenarios were implemented. For this analysis, the integrated river management (IRM) model was developed. The IRM model linked an urban wastewater system (IUWS) model with ecological models. The IUWS is a mechanistic model that incorporates the river water quality model, a wastewater treatment plant (WWTP) with activated sludge technology, and discharges from the sewage system. The ecological status of the waterways was evaluated with the ABI, which was predicted using season-specific models. In the scenario analysis, the inclusion of a new WWTP with carbon, as well as with carbon and nitrogen removal, and the addition of retention tanks before the discharges of combined sewer overflows (CSOs) were assessed. The new WWTP with carbon and nitrogen removal would bring better restoration of the ecological water quality. The retention tanks would help to enhance the ecological status of the rivers during rainy seasons.

Knowledge of the current water quality of the Cuenca River and the variables that influence its ecological status, in conjunction with the results of the proposed scenarios to restore the aquatic ecosystem, would enable stakeholders to gain insight for implementing measures to improve the ecological water quality of the Cuenca River and its tributaries. The implementation of the integrated river management (IRM) model that links an integrated urban wastewater (IUW) model with ecological models can also be replicated in other basins worldwide.

Samenvatting

Wereldwijd is de waterkwaliteit gedegradeerd, en dit hoofdzakelijk door menselijke oorzaak, zoals intensieve landbouw, industriële productie, mijnbouw, onbehandeld afvalwater en stedelijke afvoer. De lozing van vervuilende stoffen in waterwegen veroorzaakt niet alleen veranderingen in de chemische samenstelling van het water, maar tast ook de morfologische structuur van de rivierbedding aan. Bovendien worden de waterwegen beïnvloed door andere morfologische veranderingen, zoals erosie, verandering in landgebruik op en rond de oevers, en aanleg van reservoirs. Zowel verandering in de samenstelling van het water als morfologische verandering in de waterwegen veroorzaken een degradatie van de rivierecologie. Het aquatisch ecosysteem van het Cuenca Rivierbekken, gelegen in het zuiden van de Andes van Ecuador, wordt verstoord door de aanwezigheid van waterkrachtcentrales en talrijke woningen, in het bijzonder in de stad Cuenca. Lokale belanghebbenden hebben pogingen ondernomen om de ecologische status van de Cuenca rivier en zijrivieren te herstellen. Hun inspanningen waren vooral gericht op het afvalwaterbeheer. Desondanks blijft de ecologische status van de rivier aangetast door verschillende bronnen van vervuiling. De algemene doelstelling van deze thesis is de ontwikkeling van een geïntegreerd rivierbeheer (IRM) model voor ecosysteem analyse, onder huidige en mogelijke toekomstige condities, dat gebruikt kan worden als een beslissingsondersteunend systeem voor het beheer van het Cuenca Rivierbekken. Deze studie analyseert ook de fysicochemische, microbiologische en morfologische variabelen die de macro-invertebraten en biologische en ecologische waterkwaliteit van het Cuenca Riviersysteem beïnvloeden. Dit werd gedaan met behulp van geavanceerde statistische methoden en modellen. Daarenboven werden de verstoringsbronnen zoals lozingspunten, gecombineerde riooloverlopen en afvoer van afvalwaterbehandeling opgenomen in een mechanistisch geïntegreerd stedelijk afvalwater (IUWS) model. De resultaten van dit IUWS model werden gekoppeld met ecologische modellen om de waterkwaliteit in de Cuenca Rivier te beoordelen en te voorspellen, onder huidige en gesimuleerde condities.

Deze studie werd ontwikkeld voor het Cuenca Rivierbekken, met focus op twee gebieden met de hoogste verstoring van het aquatische ecosysteem. Met name de verstedelijkte gebieden van de stad Cuenca en periferie, en het Machangara sub-bekken dat twee dammen omvat voor waterkrachtproductie, werd geanalyseerd. Om een geïntegreerde waterkwaliteits beoordeling uit te voeren, werden data verzameld tijdens droge en natte seizoenen. Daarbij werd informatie over microbiologische en fysicochemische condities, evenals morfologische karakteristieken verzameld. Verder werden aquatische macro-invertebraten bemonsterd en geïdentificeerd tot op familieniveau.

In de stedelijke en substedelijke gebieden van het Cuenca Rivierbekken werd de biologische waterkwaliteit geëvalueerd in relatie tot de chemische lozing door rioolafvoer tijdens droge en natte seizoenen. De Andische Biotische Index (ABI) en de Biologische Monitoring Werkgroep aangepast aan Colombia (BMWP-Col) werden gebruikt om de biologische waterkwaliteit te beoordelen. Elk van deze biologische indices registreerde een betere biologische waterkwaliteit stroomopwaarts van de stad in vergelijking met locaties stroomafwaarts. Bovendien toonden deze indices betere condities aan tijdens het regenseizoen, gebaseerd op de aanwezigheid van meer gevoelige macro-invertebratenfamilies. Deze indices waren beter gerelateerd aan de zuurstofsaturatie dan aan de 5-daagse biologische zuurstofvraag (BOD_5), nutriënten- en chloride concentraties. De relatie tussen BOD_5 en nutriëntenconcentraties met de variatie van beide biologische indices was duidelijker in het droogseizoen. Maar op sommige plaatsen werden deze indices meer beïnvloed door morfologische aspecten dan door polluenten. De ABI index bleek meer geschikt voor de hoge Andes regio's.

In het Machangara Rivierbekken werden de abiotische voorkeuren van drie families van de Ephemeroptera—Plecoptera—Trichoptera (EPT) taxa, *Baetidae*, *Leptoceridae and Perlidae,* geëvalueerd. Met gebruik van gegeneraliseerde lineaire modellen (GLMs) werden habitat-geschiktheidsmodellen geconstrueerd, die de relatie tussen de waarschijnlijkheid van voorkomen van deze vervuilingsgevoelige macro-invertebratenfamilies en fysicochemische waterkwaliteit analyseren. Elk taxon had een verschillende predictor van zijn voorkomen, en bij gelijke predictors was de range verschillend voor elk taxon. In totaal hadden acht fysicochemische en morfologische variabelen een substantiële invloed op de uitkomst van de drie modellen.

Op vergelijkbare wijze werd de relatie tussen microbiële pathogenen en macrobenthische invertebratentaxa onderzocht in de Machangara Rivier. Habitatgeschiktheidsmodellen gebaseerd op beslissingsboommodellen (DTMs) werden gebruikt om regels te genereren die de aanwezigheid en abundantie van bepaalde benthische families linken aan microbiële pathogenen-normen. Als resultaat werden twee verschillende betrouwbare modellen verkregen als proxy indicatoren in een preliminaire beoordeling van microbiële pathogenenvervuiling in de rivieren. De bovengenoemde DTMs voorzien een onrechtstreekse, relatieve en snelle methode om na te gaan of twee Ecuadoraanse watergebruiksregulaties gerelateerd aan microbiële pathogenen vervuld zijn. Echter, deze primaire evaluatie moet gepaard gaan met analyses om te bevestigen dat microbiële waterkwaliteitsstandaarden voldaan zijn.

Biologische studies geven aan dat er substantiële verschillen zijn tussen seizoenen in tropische rivieren. Met de informatie verzameld in rivieren die door de stad en stadsrand van Cuenca stromen, werden seizoensspecifieke en seizoensoverkoepelende modellen gevormd om de biologische waterkwaliteit te voorspellen. Aangezien de ABI gebruikt werd als indicator voor biologische waterkwaliteit, werden er modellen ontwikkeld om de ABI te voorspellen met fysicochemische en morfologische variabelen als predictors. De predictors werden verkregen met gebruik van drie soorten GLMs: Gaussiaans, Gamma en Invers Gaussiaans. De seizoensspecifieke modellen waren niet accurater dan seizoensoverkoepelende modellen. Verder waren de voorspellingen van de biologische waterkwaliteit in het stedelijk gebied accurater dan de voorspellingen in referentie sites. De voornaamste variabelen die de ABI voorspellen tijdens het droog seizoen, waren BOD_5, ammoniak en orthofosfaat, terwijl opgeloste zuurstof (DO), zuurstofsaturatie (OS), nitraat en totaal aantal deeltjes belangrijk waren tijdens het regenseizoen.

Om de ecologische waterkwaliteit van de Cuenca Rivier en zijrivieren te herstellen, werden vier scenario's geïmplementeerd. Voor deze analyse werd het geïntegreerde rivierbeheer (IRM) model ontwikkeld. Het IRM model linkt een stedelijk afvalwater systeem (IUWS) model met ecologische modellen. Het IUWS is een mechanistisch model dat een rivierwaterkwaliteitsmodel, een afvalwaterbehandelingsinstallatie (WWTP) met geactiveerd slib technologie, en afvoer van het rioolsysteem incorporeert. De ecologische status van de waterwegen werd geëvalueerd met de ABI, dat voorspeld werd met seizoensspecifieke modellen. In de scenarioanalyse werd de inclusie van een nieuwe WWTP met koolstof-, evenals met koolstof- en stikstofverwijdering, en de toevoeging van retentietanks voor de afvoer van gecombineerde riooloverlopen (CSOs) beoordeeld. De nieuwe WWTP met koolstof- en stikstofverwijdering zou een beter herstel brengen van de ecologische waterkwaliteit. De retentietanks zouden helpen om de ecologische status van de rivieren in het regenseizoen te verbeteren.

Kennis van de huidige waterkwaliteit van de Cuenca Rivier en de variabelen die de ecologische status beïnvloeden, in combinatie met de resultaten van het voorgestelde scenario om aquatische ecosystemen te herstellen, kan de belanghebbenden in staat stellen om inzicht te krijgen in het implementeren van maatregelen om de ecologische waterkwaliteit van de Cuenca Rivier en zijrivieren te verbeteren. De implementatie van het geïntegreerde rivierbeheer (IRM) model dat een geïntegreerd stedelijk afvalwater (IUW) model met ecologische modellen combineert, kan ook wereldwijd in andere stroomgebieden worden toegepast.

Chapter 1. General introduction

1.1 Problem definition.

The health and well-being of a population, as well as the environment, depends on the available water quantity and quality. Worldwide, the available fresh water is affected by pollution and overexploitation. As such, water quality is mainly degraded by anthropogenic causes such as intensive agriculture, industrial production, mining, untreated wastewater and urban runoff (UN-Water 2011). Almost 80% of the wastewater in the world is discharged into water bodies without any treatment. Moreover, this percentage of untreated wastewater increases to more than 95% in some less developed countries (UN-WWAP 2017). In Latin America, the main use of water is destined for agriculture (~71%), followed by domestic supplies (~17%) and industrial use (~12%) (Aquastat 2016). In this region, around 30% of the wastewater collected in urban sewerage systems receives treatment before it is discharged into water bodies (Ballestero et al. 2015). In Ecuador, a country located in Latin America, its water usage is similar to that of the entire region, with a higher consumption for agriculture (~80%) and lower consumption for municipal (~13%) and industrial uses (~7%) (FAO 2016). Barely 7% of wastewater from its urban and industrial origins in Ecuador receives treatment before it is discharged into water bodies. Despite these facts, this natural resource receives better management in Cuenca's region in the Southern Andes of Ecuador, in which around 80% of the wastewater is treated (CEPAL 2012).

The discharges of pollutants into the waterways not only produce changes in the water's chemical composition, but discharges also affect the morphological structure of water channels (Walsh et al. 2005). Additionally, the waterways are affected by other morphological modifications such as changes of a slope by variation of the length of a river, erosion, deforestation, changes in land use and reservoirs (Galay 1983, Boix-Fayos et al. 2007). Both changes in the water composition and morphological variation in the waterways cause a disturbance of the aquatic ecosystems (De Pauw et al. 2006), producing a degradation in the river ecology. Preserving river ecology has prompted the development of policies and management plans to conserve or restore water resources. In urban ecosystems, these plans must include measures that avoid the discharge of rain and wastewater from cities that can harm the population and ecological status of the receiving water bodies (Delleur 2003, Solvi 2006).

Although efforts to improve the ecological status of the Cuenca River have been applied since 1984, with the implementation of different measures such as the expansion of the combined sewer network and the construction of the Ucubamba Wastewater Treatment Plant (U-WWTP) (ETAPA-EP 2007); the water quality in rivers continues to be affected by different contributing sources of pollution (Fernandez de Cordova and González 2012).

The ecological effects include impacts mainly on benthic invertebrates, fish and aquatic vegetation (Rahaman and Varis 2005). Benthic invertebrates, commonly referred to as macroinvertebrates, have been largely applied as bioindicators around the world to assess the water quality of water bodies (Gabriels et al. 2010). Their use as bioindicators is due to the fact that macroinvertebrates respond to the pressures from both organic pollution and morphological disturbance in waterways and their surrounding areas (Cairns and Pratt 1993). During most of their life cycle, the majority of macroinvertebrate families remain in a specific location in the aquatic system (Gabriels et al. 2010). Consequently, they can reflect a cumulative effect of chemical pollution in the aquatic biota over a period of time (Džeroski et al. 2000). The impact of habitat loss is also revealed by macroinvertebrates, a characteristic that is not detected by traditional methods used to assess the water quality (USEPA 1997). Finally, the variable degree of tolerance to the disturbance of macroinvertebrates, allows these animals to reveal the status and the quality of a water body (De Pauw et al. 2000).

To understand the response of macroinvertebrates to different stressors, powerful statistical techniques have been used. With these techniques, habitat-suitability models (HSMs) and ecological assessment models (EAMs) have been developed. The HSMs have been widely used in ecology to understand the main environmental variables that have influence over a specific species (Van Horne and Wiens 1991, Rondinini et al. 2011). With regard to aquatic environments, the HSMs are used to identify which physicochemical and morphological variables influence the presence/absence or abundance of macroinvertebrates (Mouton et al. 2009a, Lock and Goethals 2013). The EAMs link the main physicochemical and morphological variables to an index developed via bioindicators (Holguin-Gonzalez et al. 2013a). Generally, these kinds of indexes are used to assess water quality.

Pollutants enter water bodies from both point and diffuse discharges. Point discharges originate in the urban sewer networks with or without treatment. Origins are collectors from houses and industries, but stormwater can also add to this and lead to the discharge of diluted wastewater in surface waters via combined sewer overflows. While, agriculture, livestock and surface water outfalls are sources of diffuse discharges. To manage the wastewater in an urban system, the concept of an integrated design and management is of paramount importance. This design concept known as an integrated urban wastewater system (IUWS) includes the sewer network, the treatment plant and the receiving water bodies (Delleur 2003). With the help of mechanistic models that integrate the concept of an IUWS, the concentration of organic pollutants in waterways can be obtained. Furthermore, the use of the IUWS model can be applied as a planning tool, in which the performance of the wastewater system and its potential interactions and synergies can be analyzed, providing cost-effective solutions (Benedetti et al. 2013).

Although the IUWS model allows for the calculation of the concentration of organic pollutants from the sewage system in the receiving water body, these results do not provide an ecological status of a river. On the other hand, ecological models can show the physicochemical variables that can influence the ecology of water bodies. To understand the influence of an urban sewage system and its compounds in rivers, it is necessary to develop an integrated river management (IRM) model that links the IUWS with ecological models. With an IRM model, the impact of different measures to be taken in the Cuenca sewage system to improve the ecological status of the Cuenca River can be analyzed. A summarizing framework for management and restoration of an urban river is presented in Fig. 1.1.

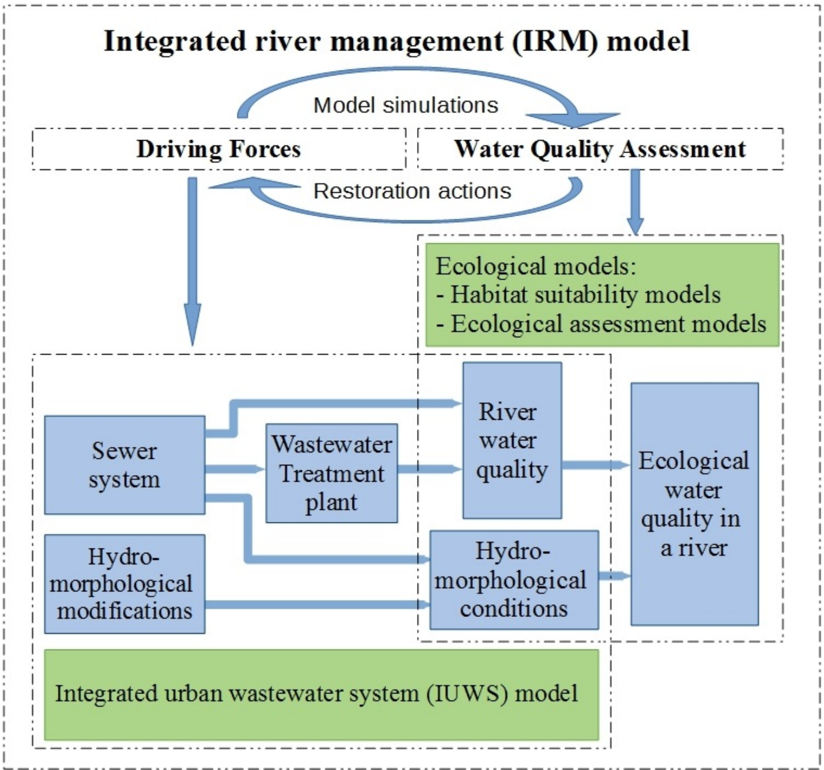


Fig. 1.1. Framework for an integrated water system model for management and restoration of an urban river.

1.2 Scope and objectives

The general aim of this dissertation is to develop an integrated river management model for ecosystem analysis - under current and new conditions - that can be used as a decision support tool in the management of the Cuenca River basin, which is located in the southern Andes of Ecuador. This study, with the help of advanced statistical methods and models, also analyzes the physicochemical, microbiological and morphological variables that could influence macroinvertebrates, and the biological and ecological water quality of the Cuenca River system. In addition, driving pressures such as outfalls, combined sewer overflows and discharges from wastewater treatment plants are included in a mechanistic integrated urban wastewater (IUWS) model. The results of this IUWS model are linked with ecological models to assess and to predict the water quality in the Cuenca River under current and simulated conditions.

The results of this Ph.D. research can be applied to improve the water quality of the Cuenca River system. Furthermore, a base line of current conditions during dry and rainy seasons is provided. This baseline can be used to compare its variation with the implementation of different measures that can be taken to improve the water quality. In addition, the benefits of a set of possible measures to enhance the water quality in the Cuenca River are analyzed. This research provides a framework in water quality assessment and in the development of statistical and mechanistic models that can be replicated in other basins around the world.

The research questions which this dissertation addresses are the following:

1. What is the current biological water quality in the urban and suburban areas of the Cuenca River? What is the impact of the sewage outfalls in the urbanized area of the Cuenca River basin? Which factors influence the water quality in the urban and suburban areas of the Cuenca River basin? Is the biological water quality influenced by different variables in each season? Which biological index is more suitable for the high Andean region? (cf. Chapter 3).
2. What is the water quality in the Machangara River basin, a sub-basin of the Cuenca River? Which physicochemical and morphological variables are related to the occurrence of sensitive taxa - the Ephemeroptera—Plecoptera—Trichoptera (EPT) taxa? How were the sensitive taxa affected by different physicochemical variables, affected including those with different values of these variables? (cf. Chapter 4).
3. What is the current microbial water quality in the Machangara River basin? Is the biological index, BMWP-Col, influenced by microbial pathogens? Can the presence or abundance of sensitive taxa of macroinvertebrates be influenced by microbiological water quality? Can sensitive macroinvertebrates be used as proxy indicators in the fulfillment of the microbiological water quality regulations? (cf. Chapter 5).
4. How well do ecological models perform in one season versus another season? What is the benefit of season-overarching models? Is the use of a selective combination of season-specific models more effective than the use of a season-overarching model? (cf. Chapter 6).
5. Can various measurements be simulated in an integrated urban wastewater system? Which measures are more effective to improve the current water quality in the urban and suburban areas of the Cuenca River basin? Does the use of retention tanks located before the discharges of the combined sewer overflows influence the water quality of the rivers? What is the benefit of a wastewater treatment plant (WWTP) with carbon removal in comparison with a WWTP with

carbon and nitrogen removal in the water quality of the Cuenca River? (cf. Chapter 7).

In order to accomplish the general objectives of this doctoral research project, each chapter answers the specific research questions previously described. Fig. 1.2 shows the organization of the thesis, and the link between its different chapters.

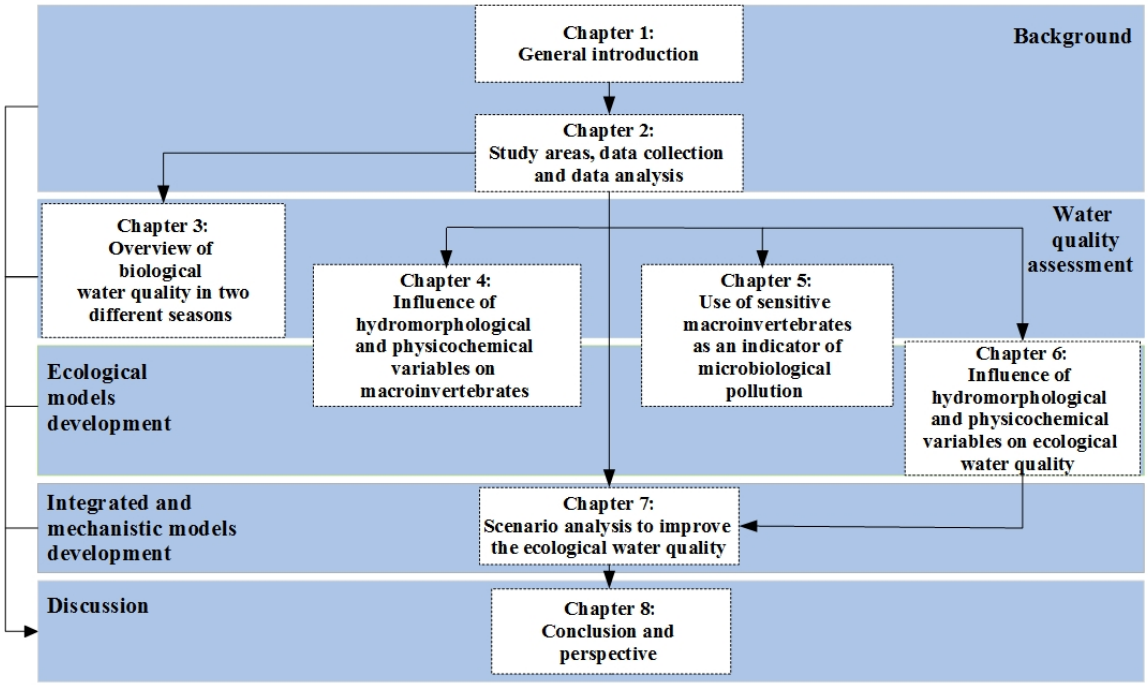


Fig. 1.2. Schematic diagram showing the organization and the link between the different chapters within the thesis.

Chapter 1 gives a general introduction to the relevance of this study. This chapter also includes the scope, objectives and the research hypothesis questions of this thesis.

Chapter 2 provides an overview of the study areas and the monitoring methods applied to the data collection. Additionally, an overview of the statistical methods applied in the data analysis and in the construction of the ecological and biological models is included. Finally, a brief description of the integration of ecological models with river models is presented.

Chapter 3 assesses the biological water quality and identifies the impact of the sewage outfall on the rivers in the urban and suburban areas of the Cuenca River basin and its affluents. This assessment was performed during both dry and rainy seasons, utilizing two biological indexes: the Andean Biotic Index (ABI) and the Biological Monitoring Working Party adapted to Colombia (BMWP-Col). The changes of these biological indexes were analyzed with the variation of physicochemical and morphological variables during both the dry and rainy seasons.

Chapter 4 evaluates the water quality in the Machangara River basin, an affluent of the Cuenca River. Moreover, this chapter analyzes the abiotic preferences of three families of the Ephemeroptera—Plecoptera—Trichoptera (EPT). This study was done using statistical models to develop habitat-suitability models, in which the probability of occurrence of these pollution-sensitive macroinvertebrates families was related to the physicochemical water quality conditions.

Chapter 5 develops habitat-suitability models that relate families of macroinvertebrates with microbial pathogens for fast detection of unsuitability of water. Furthermore, the possible use of these models as indirect, approximate and rapid methods of checking the fulfillment of Ecuadorian regulations for water use related to microbial pathogens is analyzed. This analysis was performed in the Machangara River basin.

Chapter 6 presents an analysis of the use of season–specific models and season-overarching models to predict the biological water quality in the Tomebamba and Cuenca Rivers. These models were classified as ecological assessment models (EAMs) and forecast the Andean Biotic Index (ABI), which is an indicator of biological water quality, with physicochemical and morphological variables as predictors. This study was performed in the urban and suburban areas of the Cuenca River basin.

Chapter 7 implements a scenario analysis to restore the ecological water quality in the Cuenca River and its tributaries. For this analysis, an integrated urban wastewater system (IUWS) model was developed and linked with the ecological models developed in Chapter 6. The IUWS is a mechanistic model that incorporates the river water quality model, a wastewater treatment plant (WWTP) with activated sludge technology, and discharges from the sewage system. In the analysis, four scenarios that would enhance the current ecological water quality were tested and evaluated.

Chapter 8 includes a general conclusion of this dissertation, as well as some recommendations for further research provided.

Chapter 2. Materials and methods

2.1 Study area

Two areas of the Cuenca River basin were considered for this research: the urban and suburban areas and the Machangara River basin, a sub-catchment of the Cuenca River basin.

2.1.1 Urban and suburban area of the Cuenca River basin

The first 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 an Andean mountain stream that is part of the Paute Upper basin and the Hydrographic Demarcation Santiago, one of the Amazon tributaries. The Cuenca River is located downstream from the city of Cuenca (Fig. 2.1), 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 (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 River 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 December, while the rest of the year constitutes the dry season (Fig. A1)(Aereopuerto-Mariscal-Lamar 2012). The average flow of the Cuenca River measured before its discharge into the Paute River was 28 $m^3 \cdot s^{-1}$, and the average minimum flow was 11 $m^3 \cdot s^{-1}$, which occurs between August and September (Cordero Domínguez 2013).

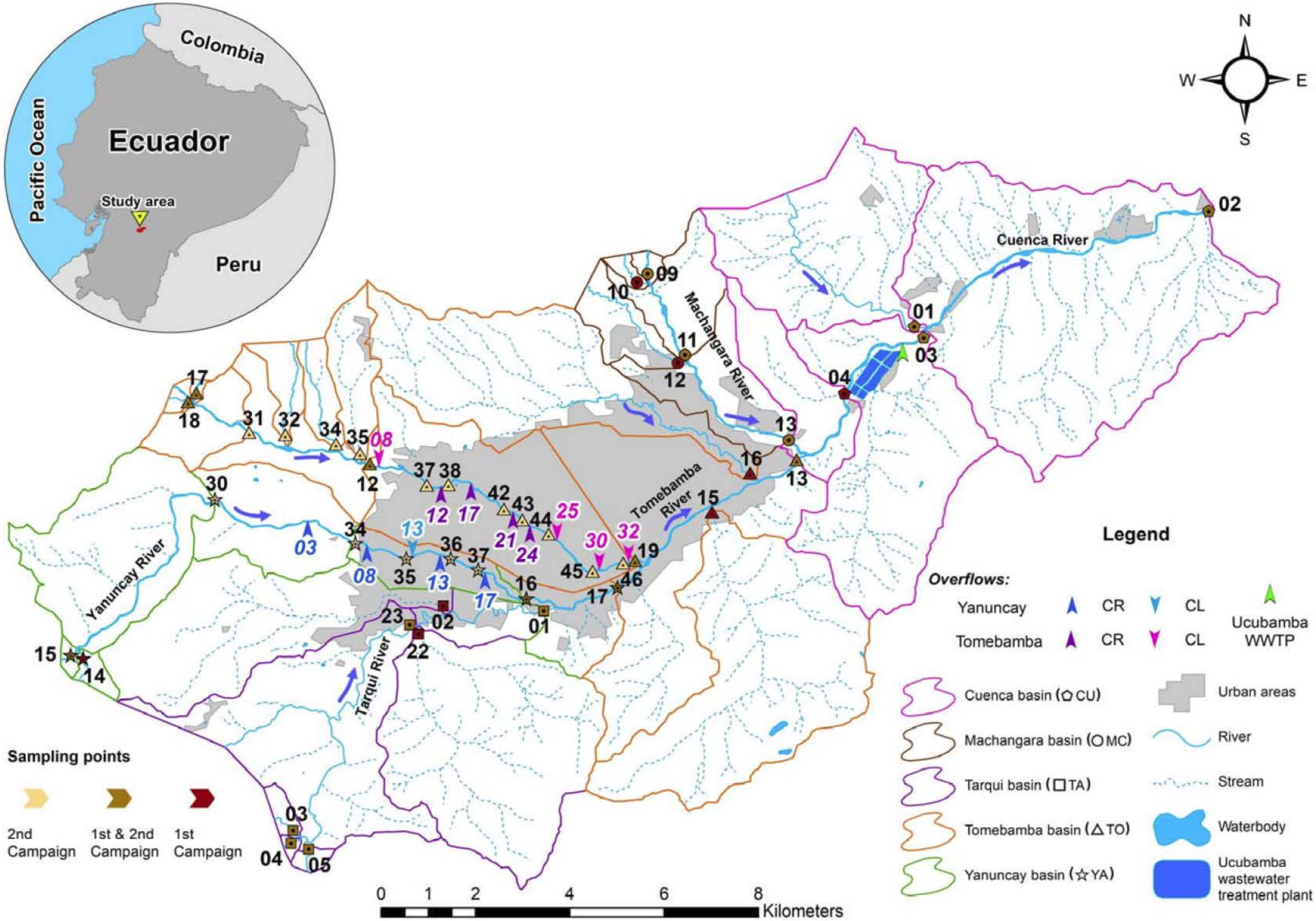


Fig. 2.1. Map with rivers (and sampling sites) and sewer outfalls (sewer and overflow discharges) from the Cuenca River basin.

Water is primarily extracted from the Yanuncay, Tomebamba and Machangara subcatchments for drinking water for Cuenca and the surrounding areas. To a lesser extent, water from these rivers is used to supply industry and livestock as well as irrigation for agriculture, with the exception of the Tarqui River basin, in which water is mainly used for agricultural irrigation and livestock production.

2.1.2 Machangara River basin

The Machangara River is an Andean mountain river, which is an affluent of the Cuenca River. It is about 37 km in length and at the end of its path, crosses the city of Cuenca (Fig. 2.2).

The area of Machangara River basin is about 325 km^2, of which 79.1% is forest protected by the Ecuadorian Government, 8.9% is a mosaic between urban area, pastures and crops, 7.1% is used as pastures, 2.4% is urban area, 1.1% is water bodies and 1.4% is bare soil (Fig. 2.2A). The altitude of the basin varies from 2440 to 4420 m a.s.l. and its mean altitude is 3557 m a.s.l. The average annual rainfall in the basin varies between 877 mm and 1363 mm per year, while the average annual temperature fluctuates between 16.3 °C and 9.0 °C in the lower and upper areas respectively (PROMAS-UCuenca 2010, Aereopuerto-Mariscal-Lamar 2012). The average flow of the Machangara River measured from 1964 to 2010, before discharging into the Tomebamba River, was 8.4 $m^3 \cdot s^{-1}$, the average minimum was 5.3 $m^3 \cdot s^{-1}$ in August and the average maximum was 14.6 $m^3 \cdot s^{-1}$ in May (Estrella and Tobar 2013), cf. Fig. A2 in Appendix A for monthly averages.

2.1.3 Improvements developed to enhance the water quality in the Cuenca River system

The Water Supply and Wastewater Management Municipal Company, ETAPA-EP, has been working since 1984 to improve the water quality in the rivers that pass through the city of Cuenca. Accordingly, projects have been implemented in Cuenca and its surrounding areas, including the construction of the Ucubamba Wastewater Treatment Plant (U-WWTP), the expansion of sewer interceptors and the enlargement of the combined sewer network (ETAPA-EP 2007). The U-WWTP treats both municipal and industrial wastewater so that its daily-average effluent meets the Ecuadorian standards (MAE-Ecuador 2015). These standards regulate the thresholds that are discharged by the effluent of a WWTP into a freshwater body. However, water quality in rivers is being affected by different contributing sources of pollution such as point discharges from sewer networks and industries (Fernandez de Cordova and González 2012), storm water overflows from the combined sewer network and surface water outfall (SWO) (Hvitved-Jacobsen 1982, Mulliss et al. 1997, Weyrauch et al. 2010, Passerat et al. 2011) and excess flow from the by-pass located before the U-WWTP (Jerves-Cobo et al. 2018b). Adding to the degradation of the water conditions in the rivers are the diffuse fluxes of organic pollution from extensive livestock use in the Tarqui River basin, which is less widespread in the other catchments (Beltrán et al. 2013).

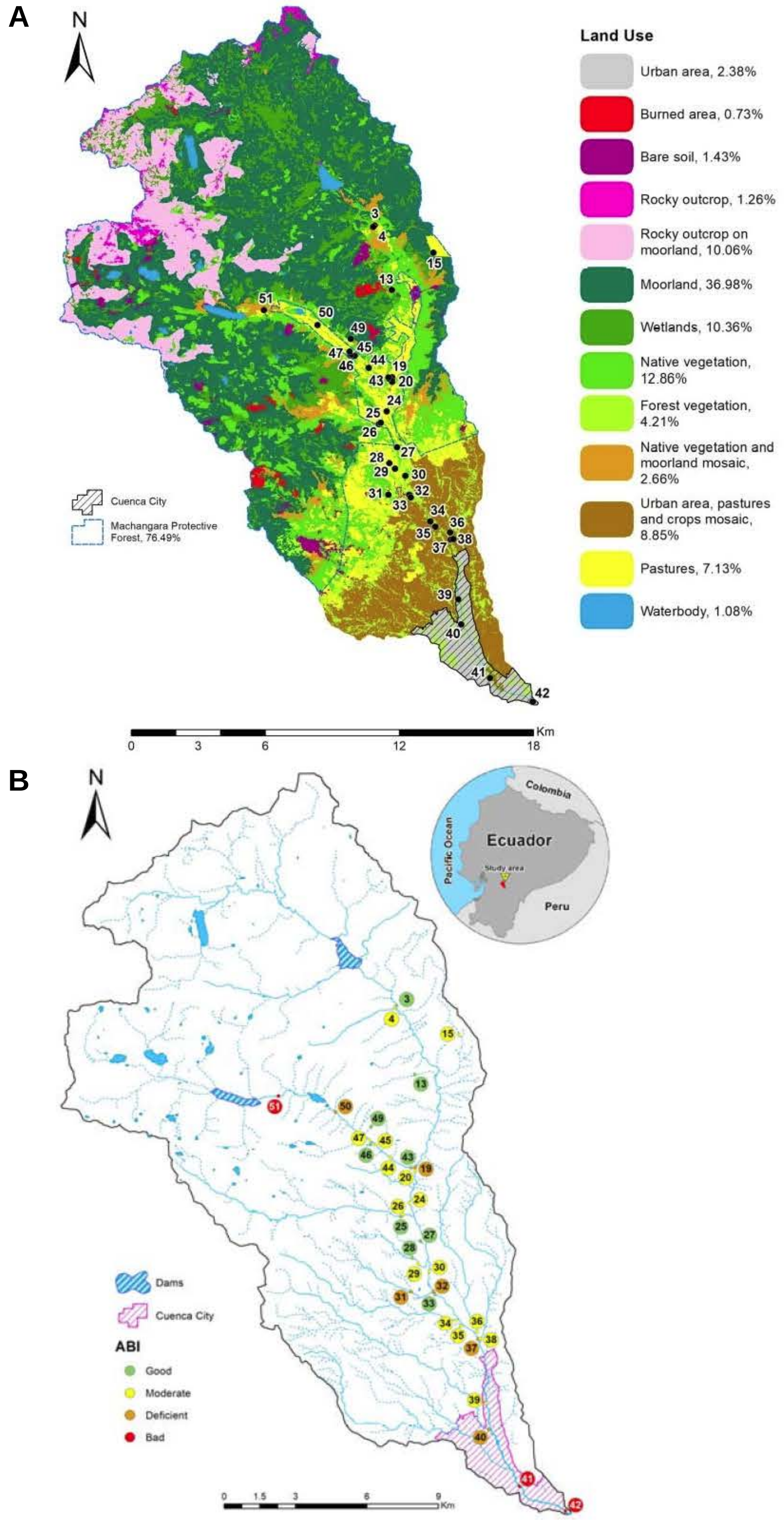


Fig. 2.2. Map with sampled site locations in the Machangara River basin: (**A**) with the land use; (**B**) with the qualifications given by the Biological Monitoring Working Party score adapted to Colombia (BMWP-Col).

2

2.2 Data Collection

2.2.1 Urban and suburban area of the Cuenca River basin

In the urban and suburban areas of the Cuenca River basin, 43 sites were sampled (Fig. 2.1), which were located in Cuenca and nearby areas. Of these sites, 27 were sampled during the dry season of July 2015, while during the rainy season of March 2016 the samples were taken from 35 sites. Nineteen sites were sampled during both seasons. The samples were collected during the day, from 8:00 am to 5:00 pm. At each sampling site, environmental (physical-chemical and hydro-morphological) and biological (macroinvertebrates) conditions were measured. The sampling sites were selected based on two criteria: macro-locations and micro-locations (Strobl and Robillard 2008). The first criterion was to understand the general condition 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 assess local impacts in the rivers, micro-locations were selected upstream and downstream from the outfalls (Ward 1973, Strobl and Robillard 2008). Thus, the highest sampling sites upstream in the four main rivers were located prior to the start of the city's suburban area. Other sampling sites were placed in the tributaries, upstream from the connection to the main river, the Tomebamba. Additionally, specific sites in the rivers upstream and downstream from the main direct discharges or overflows from the sewage system, and ahead of and below the discharges of U-WWTP were measured. Similarly, samples were taken from the Tomebamba and Yanuncay Rivers before and after the main overflows from the sewage system during the rainy season.

In each sampling site, 28 physicochemical, hydraulic and microbiological variables were measured. Six of these variables were registered 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 sub-flows computed in each subsection. The rest of the parameters shown in Table 2.1 were measured in the laboratory of Sanitation at the University of Cuenca. The physicochemical information of the Cuenca River was complemented with nine samples taken before and after the discharge of the U-WWTP during 2017 (Espinoza Berrezueta and Zumba López 2018). Additionally, in each location and its surroundings, information was compiled on elevation, land use, river morphology, substrate characteristics, macrophytes, shading in the rivers and flow variation, according to the field protocol shown in Table A1 in Appendix A. This protocol was modified based on the physical assessment protocol called the Australian River Assessment System (AusRivAS) (Parsons et al. 2002), and the United Kingdom and the Isle of Man River Habitat Survey (RHS) (Raven 1998).

Table 2.1. Summary of the physical, chemical and microbiological data collected in the studied area of the Cuenca River basin in Ecuador 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 (NH_4)	$mg\cdot L^{-1}$	1.0	±	4.0	BDL	26.4	0.1	2.1	0.2
Nitrate-N (NO_3)	$mg\cdot L^{-1}$	0.3	±	0.2	BDL	1.7	0.3	0.3	0.3
Nitrite-N (NO_2) [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	$MPN.100mL^{-1}$	4.1E+05	±	3.5E+01	1.4E+03	4.3E+10	1.9E+05	8.4E+05	2.3E+05
	$CFU.100mL^{-1}$	7.5E+04	±	1.7E+01	3.9E+02	8.4E+09	5.6E+04	1.2E+05	5.4E+04
Fecal coliforms	$MPN.100mL^{-1}$	1.3E+05	±	3.1E+01	4.9E+02	9.2E+09	8.6E+04	2.0E+05	9.4E+04
	$CFU.100mL^{-1}$	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 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

2.2.2 *Machangara River basin*

In the Machangara River basin, the second area of study of this research, 33 sites were sampled in 2012 (Fig. 2.2), sites whose altitudes varied from 2451 to 3428 m a.s.l. The locations were chosen along the catchment according to land use (Fig. 2.2A). The dataset used in the Machangara River basin was collected and measured once during the rainy season in February and March 2012, while the validation data set was sampled in the last half of July 2015 in the dry season and in March of 2016 in the rainy season. In the validation datasets, the samples were collected from 14 points in July of 2015 and from 11 sites in March of 2016 (Fig. A3 in Appendix A). In each point sampled in 2012, 17 physicochemical, hydraulic, microbiological and biological variables were measured (Table 2.2). From this data, four variables were measured in situ: water temperature, conductivity, dissolved oxygen (DO) and pH with an ORION 5Star 1219001 (Thermo Scientific, Waltham, MA, USA) multi-parameter probe. Flow velocity was measured using the float method described by the U.S. Environmental Protection Agency (Dohner et al. 1997). The rest of the parameters and the methods used by their determination in the laboratory of Sanitation at the Water Supply and Wastewater Management Municipal Company ETAPA—EP in Ecuador are shown in Table 2.2.

Table 2.2. Summary of the physical, chemical and microbiological data collected in the Machangara River basin in Ecuador based on 33 samples in 2012.

Parameter	Units	Mean Value		Standard Deviation	Min Value	Max Value	Median Value
Mean depth	m	0.3	±	0.3	0.0	1.6	0.3
Flow velocity	$m \cdot s^{-1}$	0.6	±	0.4	0.1	1.8	0.5
Temperature	°C	11.5	±	1.1	9.1	13.4	11.9
pH		7.6	±	0.5	6.3	8.4	7.7
Dissolved oxygen (DO)	$mg \cdot L^{-1}$	9.1	±	1.5	6.7	12.6	9.5
Total solids (TSol)	$mg \cdot L^{-1}$	89.1	±	51.7	19.0	190.0	74.0
Turbidity	NTU	7.7	±	11.1	0.5	48.2	3.7
True color (color)	HU	14.4	±	8.5	0.0	40.0	14.0
Specific conductivity	$\mu S \cdot cm^{-1}$	91.6	±	44.1	13.2	238.0	82.3
Phosphates	$mg \cdot L^{-1}$	0.1	±	0.1	0.0	0.6	0.0
Nitrate + Nitrite-N	$mg \cdot L^{-1}$	0.1	±	0.1	BDL	0.7	0.0
Ammonium nitrate	$mg \cdot L^{-1}$	0.0	±	0.1	0.0	0.4	0.0
Organic nitrogen	$mg \cdot L^{-1}$	0.6	±	1.2	0.0	6.6	0.1
Biochemical oxygen demand 5 day (BOD_5)	$mg \cdot L^{-1}$	1.1	±	2.4	BDL	13.0	0.4
Chemical oxygen demand (COD)	$mg \cdot L^{-1}$	9.9	±	8.4	2.0	46.0	8.0
Fecal coliforms	MPN.100 mL^{-1}	3.6E+04	±	1.0E+05	4.5	5.4E+05	79

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
BDL = Below Detection Limit

2.2.3 Macroinvertebrates

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 x 0.30 m and mesh size of 300-500 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. Live animals were sorted in a white tray. Macroinvertebrates were identified up to family level with the help of a stereomicroscope (Roldán Pérez 1988, Álvarez 2005, Encalada et al. 2011).

During both seasons (July of 2015 and March of 2016), in the urban and suburban areas of the Cuenca River basin, 43 taxa of macroinvertebrates were identified (Table

A2 and Table A3 in Appendix A). In the Machangara River, during February and March of 2012, 37 families were collected (Table A4 and Table A5), while during July of 2015 and March of 2016, 37 families were identified (Table A6 in Appendix A).

At each sampling site in the urban and suburban areas of the Cuenca River basin, in order to assess the water quality two biological indexes were calculated: the ABI, which was built based on adaptations of the BMWP to South America (Encalada et al. 2011, Ríos-Touma et al. 2014) and the BMWP adapted to Colombia (BMWP-Col) (Roldán Pérez 1988, 2003, Álvarez 2005). For samples taken in the Machangara River basin, only the BMWP-Col was calculated. For the determination of the BMWP-Col *Acari*, *Dytiscidae*, *Lumbricidae* and *Stenopsychidae* were omitted and to obtain the ABI, *Dugesiidae*, *Glossiphoniidae*, *Lumbricidae* and *Stenopsychidae* were excluded (Table A2 and Table A3 in Appendix A). The ABI and the BMWP-Col give a five-water quality classification in function of the sum of sensitivity scores obtained in each location: bad (≤15), deficient (16–35), moderate (36–60), good (61–99), and very good (>99) (Álvarez 2005, Zúñiga et al. 2009). The sensitivity score of these biological indexes ranges from one for very tolerant taxa to 10 for the most sensitive families. The families that did not appear as part of the biological indexes were excluded from this calculation.

2.3 Statistical modelling

Two statistical techniques were used to link the presence/absence or abundance of macroinvertebrates in rivers as well as the water biological quality with variables. These variables were related to the physicochemical and microbiological water quality conditions and morphological characteristics of the rivers. The first of the techniques implemented were generalized linear models (GLMs), which were applied to obtain the HSMs and the ecological assessment models (EAMs). The HSMs predicted the presence or absence of three pollution-sensitive macroinvertebrates according to the concentration of physicochemical variables in water. The EAMs linked the biological water class based on the Andean Biotic Index (ABI), with the concentration of water physicochemical variables and morphological characteristics of rivers, in both the dry and rainy seasons. The second of the techniques implemented were decision tree models (DTMs), which were developed to obtain habitat-suitability models (HSMs) that associated the presence/absence or abundance of pollution-sensitive taxa of macroinvertebrates with fecal coliforms (FCs) concentrations. The threshold of the FCs was set according to the values established in three Ecuadorian regulations related to water use. An overview of the statistical techniques applied in the different chapters can be found in Table 2.3.

Table 2.3. Summary of statistical techniques with their input and output variables applied in the construction of the ecological or biological models

Statistical technique	Ecological or biological model	Explanatory variable	Output (Response variable) variable
Generalized linear model (GLM) with binomial distribution (Logistic regression models) – Chapter 4.	Habitat-suitability models.	BOD_5, COD, conductivity, flow velocity, fecal coliforms in logarithmic scale, pH, temperature and true color (Color)	Presence/absence of *Baetidae* (Ephemeroptera), *Leptoceridae* (Trichoptera) and *Perlidae* (Plecoptera).
Decision tree models (DTMs) – Chapter 5.	Habitat-suitability models.	Presence/absence or abundance of macroinvertebrates taxa (*Baetidae, Scirtidae, Perlidae*).	Fecal coliforms (FCs) concentrations given by the values of the fulfillment of three Ecuadorian regulations related to water use.
Generalized linear model (GLM) with Gaussian, Gamma and Inverse Gaussian adjustments – Chapter 6.	Ecological assessment models (EAMs).	*Dry season*: Nitrate, ammonium, BOD_5 and orthophosphates. *Rainy season*: Nitrate, nitrite, dissolved oxygen, oxygen saturation, total solids and river bank material.	Andean Biotic Index (ABI) Value – ABI Class

2.3.1 Generalized linear models to predict the presence or absence of families of EPT taxa.

Generalized linear models (GLMs) with binomial distribution were applied to obtain habitat-suitability models (HSMs) that predicted the presence or absence of the three sensitive EPT taxa. The taxa analyzed were *Baetidae*, *Leptoceridae* and *Perlidae* (Table 2.3). The presence or absence of a species, in connection with explanatory variables such as environmental conditions, has been frequently modeled as a binary response using the GLM (Zuur et al. 2009, Holguin-Gonzalez et al. 2013b, Everaert et al. 2014). GLMs 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, Nguyen et al. 2015, Forio et al. 2018). The GLMs were developed in order to conceptualize the environmental preferences of macroinvertebrates. In a Logistic regression, the distribution of the response variable Y_i also expresses the mean and variance of Y_i. As a result, the expected value and variance of Y_i are given by $E(Y_i) = \pi_i$ and $var(Y_i) = \pi_i \times (1 - \pi_i)$. Where, π_i signifies the probability that an analyzed taxon is present in an observation *i* and $(1 - \pi_i)$ is the probability that it is absent. The models were composed using Equation (2.1), in which X_{ni} means the explanatory variables, while α and β_j are regression parameters. All parameters in the Logistic regression models were estimated by means of maximum likelihood (Zuur et al. 2009).

$$E(Y_i) = \pi_i = \frac{e^{(\alpha+\beta_1\times X_{1i}+\beta_2\times X_{2i}+\cdots\ldots+\beta_n\times X_{ni})}}{1+e^{(\alpha+\beta_1\times X_{1i}+\beta_2\times X_{2i}+\cdots\ldots+\beta_n\times X_{ni})}} \quad (2.1)$$

The goodness-of-fit of the GLMs with binomial distribution was assessed with the Akaike information criterion (AIC) and the pseudo R^2. The AIC is a statistical performance criterion that is a trade-off between model complexity and model accuracy criteria. Thus, the best model, which has the lowest value of AIC, tends to fit closest to reality (Agresti and Kateri 2011). The pseudo R^2 is a likelihood ratio index that is analogous with the R^2, which is used in multiple linear regression techniques (Hu et al. 2006). According to this author, pseudo R^2 is defined as the proportion of the variance of the response variable that is explained by the explanatory variable. Its value gives a measurement of how well-observed outcomes are replicated by the model. Thus, when its value is closer to one, the model replicates in an excellent way the observed outcomes, while the contrary happens when the value of pseudo R^2 is close to zero. This value is calculated based on the Kullback-Leibler divergence (Cameron and Windmeijer 1997).

To identify which explanatory variables have an influence on the presence/absence of a taxon, 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). When the value of AIC 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). More details on the application of this technique and the construction of the models can be found in section 3.2.

2.3.2 Decision tree models

The explanatory variables of the DTMs were the presence/absence or abundance of three families of Ephemeroptera – Plecoptera - Trichoptera (EPT) of macroinvertebrates taxa. These families of the EPT taxa were *Baetidae, Scirtidae* and *Perlidae*. The response variables were in fulfillment of the three microbial water quality standards for fecal coliforms (Table 2.3). The DTMs are hierarchical structures, where internal nodes contain a test on the input explanatory variables. Each branch of an internal test corresponds to an outcome of the test and the prediction for the values of the response variable is stored in a leaf. Each leaf of the decision tree contains a prediction for the response variable. The decision trees explain the variation in response variables by splitting explanatory variables at certain thresholds in the node of the tree. Furthermore, each division or level can produce more nodes with branches that follow a new ordering instruction (Quinlan 1986, Quinlan 1987, Everaert et al. 2011, Lior 2014). Decision trees have been applied in numerous ecological studies such as the macroinvertebrate habitat suitability analysis (Dakou et al. 2007, Hoang et al. 2010b); because the DTM combines a reliable classification with a transparent set of rules (Everaert et al. 2011).

The accuracy of the DTMs was assessed with three measures obtained from the confusion matrix: the number of correctly classified instances (CCI), the Cohen's

Kappa statistic (κ) and the overall confusion entropy of a confusion matrix (CEN). 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 (TP + TN) divided by the sum of all values (TP + FP + TN + FN) (Kohavi and Provost 1998). The value of the CCI is expressed in percentage and ranges from 0 to 100%, where a value of 100% means that the accuracy of the model is the greatest (Fukuda et al. 2011). The Cohen's Kappa statistic (κ) (Eq. 2.2) 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, Dakou et al. 2007, Jerves-Cobo et al. 2018a). This coefficient is calculated as:

$$\kappa = \frac{(TP+TN) - \frac{(TP+FN)(TP+FP)+(FP+TN)(FN+TN)}{n}}{n - \frac{(TP+FN)(TP+FP)+(FP+TN)(FN+TN)}{n}} \qquad (2.2)$$

The interpretation of the model fit with respect to different Kappa statistic (κ) values is as follows: Poor (<0), Slight (0–0.20), Fair (0.21–0.40), Moderate (0.41–0.60), Substantial (0.61–0.80) and Almost Perfect (0.81–1.0) (Landis and Koch 1977). Models are considered good when the Kappa statistics is higher than 0.4 and the CCI is at least 70% (Goethals 2005). Finally, the CEN evaluates the confusion level of the class distribution of misclassified samples. Deeper details of the application of this technique and of the model construction can be checked in Chapter 5.

2.3.3 Ecological Model - GLM model: relation between chemical water quality and Andean Biotic Index (ABI)

Ecological assessment models (EAMs) were developed in order to establish the variables that could influence the Andean Biotic Index (ABI). These EAMs were obtained with the application of a GLM technique applied to the data collected during both the dry and rainy seasons (Table 2.3).

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 GLM 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 Equations 2.3, 2.4 and 2.5, respectively; where $E(Y_i)$, also expressed as μ_i, is the mean of the response (or dependent) variable Y, given the explanatory 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. All parameters in the GLMs were estimated by means of maximum likelihood (Zuur et al. 2009, Lovison et al. 2011, Hardin and Hilbe 2013).

Gaussian GLM:

$$\mathrm{E}(Y_i) = \mu_i = \eta_i = \alpha + \beta_1 \times X_{1i} + \beta_2 \times X_{2i} + \cdots \ldots + \beta_n \times X_{ni} \tag{2.3}$$

Gamma GLM:

$$\mathrm{E}(Y_i) = \mu_i = \frac{1}{\eta_i} = \frac{1}{\alpha + \beta_1 \times X_{1i} + \beta_2 \times X_{2i} + \cdots \ldots + \beta_n \times X_{ni}} \tag{2.4}$$

Inverse Gaussian GLM:

$$\mathrm{E}(Y_i) = \mu_i = \frac{1}{\sqrt{\eta_i}} = \frac{1}{\sqrt{\alpha + \beta_1 \times X_{1i} + \beta_2 \times X_{2i} + \cdots \ldots + \beta_n \times X_{ni}}} \tag{2.5}$$

The three-fold cross-validation procedure was implemented during the construction and validation of the GLMs (Goethals 2005). For this, each dataset was randomly stratified according to the response variable (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. 2013b, Forio et al. 2016). Each training set was fitted with the three different types of GLMs indicated in Equations 2.3, 2.4 and 2.5.

The selection of the variables that have influence on the ABI were selected with a stepwise model selection procedure that was described in Section 2.3.1.

The model's goodness-of-fit was preliminarily assessed with the pseudo R^2. In order to choose the best models, the predicted ABI values obtained with the GLMs with the highest pseudo R^2 were transformed into a categorical variable based on the water quality classification as elaborated in section 2.2.3. 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). These measures of accuracy were described in section 2.3.2. 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). 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. For more details of the construction and the selection of the ecological models, the reader is referred to section 6.2.

2

2.4 Integrated model

To predict the water quality in the Cuenca River system in current conditions and under different scenarios, two kinds of models were linked: an integrated urban wastewater system (IUWS) model and ecological assessment models (EAMs). The IUWS model is a mechanistic model that includes a sewage system, a wastewater treatment plant (WWTP) with an activate sludge process and a river water quality model. The IUWS model was developed in an open model structure simulator called WEST®. The River Water Quality Model No.1 (RWQM1) (Reichert et al. 2001) with the Activated Sludge Model No. 1 (ASM1) (Henze et al. 2000) and with the combined sewer overflows were coupled and simulated in the WEST®. The RWQM1 is based on the concept of a cascade of continuously stirred tank reactors in series (CSTRs) to represent the transport of pollutants along rivers. The ASM1 simulates carbon oxidation in aerobic and anoxic conditions as well as nitrification and denitrification processes in a WWTP. The IUWS model was constructed, calibrated and validated for both the dry and rainy seasons. The EAMs are ecological models that were described in Section 2.3.3. The ecological models predicted the Andean Biotic Index (ABI) with the resulting values of the physicochemical variables obtained in the IUWS model and the hydro-morphological characteristics observed in the Tomebamba and Cuenca Rivers. A detailed description of the integration of the models can be found in Chapter 7.

Chapter 3. Biological impact assessment of sewage outfalls in the urbanized area of the Cuenca River basin (Ecuador) in two different seasons

3

Adapted from:

Rubén Jerves-Cobo, Koen Lock, Jana Van Butsel, Guilermina Pauta, Félipe Cisneros, Ingmar Nopens and Peter L. M. Goethals. (2018). Biological impact assessment of sewage outfalls in the urbanized area of the Cuenca River basin (Ecuador) in two different seasons. *Limnologica*, 71, 8-28. doi.org/10.1016/j.limno.2018.05.003. ISSN 0075-9511

3

Abstract

In this chapter is evaluated the biological water quality in relation to chemicals discharged through sewage outfall during both dry and rainy season. The lowland area of the Cuenca River basin in the southern Andes of Ecuador, including the city of Cuenca, constituted the study area. To perform an integrated water quality assessment, data were collected of macroinvertebrates, physicochemical conditions and morphological characteristics in 43 sites in the Cuenca River and its tributaries. The Andean Biotic Index (ABI) and the Biological Monitoring Working Party adapted to Colombia (BMWP-Col) were used to evaluate the biological water quality. Both biological indexes were higher upstream than downstream from the city. Moreover, these indexes indicated better conditions during the rainy season than in the dry season, based on the presence of more sensitive families. The biological indexes related more to the oxygen saturation than to the five-day biological oxygen demand (BOD_5), nutrients and chloride concentrations. The relationship between BOD_5 and nutrient concentrations with the variation of both biological indexes was clearer in the dry season than in the rainy season. However, in some sites, these indexes were influenced more by morphological aspects than by pollutants. Both biological indexes showed similar patterns along the rivers, generally the BMWP-Col scored higher than the ABI index. The latter index was shown to be more suitable for the high Andes region as an indicator of water quality. These results could be used to monitor the implementation of river restoration actions, such as determining priorities for splitting sewer and precipitation water transport systems and needs for improved wastewater treatment facilities in specific locations.

3.1 Introduction

Urbanization and human activities such as agricultural and livestock farming increase the pressure and alteration of the natural ecosystems, introducing hydro-morphological changes and physicochemical pollution (Holguin-Gonzalez et al. 2014), which affect the living communities in water bodies (Helson and Williams 2013). Benthic macroinvertebrates are used to evaluate the water quality over time, because they respond to pressures in a river due to both physicochemical changes and hydro-morphological variations (De Pauw et al. 2006), while water quality assessment using physicochemical samples provides limited information at a specific point in time (Džeroski et al. 2000). The Biological Monitoring Working Party (BMWP), an index initially used to assess the water quality in England, which has been adapted to tropical countries (Junqueira and Campos 1998, Mustow 2002, Roldán Pérez 2003), was developed on the basis of the sensitivity to organic pollution of taxa found in water bodies. The aforementioned index was adapted to the environmental conditions of Colombia and became known as the BMWP-Col, which was used in different regions of Colombia and Ecuador (Roldán Pérez 2003, Álvarez 2005, Dominguez-Granda et al. 2011). The BMWP-Col has recently been updated with reviewed tolerance scores (TS) of pollution for taxa, according to the environmental conditions of the high Andes of Ecuador and Peru. This updated index is called the Andean Biotic Index (Ríos-Touma et al. 2014). However, the accuracy of the BMWP-Col and the ABI to assess the biological quality in water bodies in the high Andes has not yet been analyzed.

In the Ecuadorian Andes, particularly in the northern region, only few studies have been conducted to understand the interrelation between macroinvertebrates and physicochemical and morphological stressors (Jacobsen 1998, Jacobsen et al. 2003, Jacobsen 2008, Ríos-Touma et al. 2011). These authors found that key variables such as oxygen saturation, flow velocity and altitude influenced the presence of macroinvertebrate taxa. Moreover, they noticed that the impacts from the biological oxygen demand five (BOD_5) and nutrients were more visible during dry season (Jacobsen 1998). Although various monitoring efforts were developed from 1997 to 2013 in the Cuenca River basin by the Water Supply and Wastewater Management Municipal Company (ETAPA-EP), these evaluations did not follow a well-standardized methodology, leaving the obtained results incomparable (Acosta and Hampel 2015).

Three studies were carried out to determine the effect of water quality on macroinvertebrates in the Cuenca River basin. The first study, executed by Holguin-Gonzalez et al. (2013a), modeled the presence of *Trichoptera* and *Physidae* in function of the five-day biological oxygen demand (BOD_5) and fecal coliforms (FC), respectively. In the second study, Jerves-Cobo et al. (2017), found a relation between the presence of *Baetidae, Leptoceridae and Perlidae* with eight physicochemical variables in the Machangara River, a tributary of the Cuenca River. In the third study, Jerves-Cobo et al. (2018a) analyzed the relationship between macroinvertebrates and FC in the aforementioned tributary. However, neither the variation between seasons nor a completed impact of the outfalls was evaluated.

The aim of this chapter is to evaluate the biological water quality in relation to the intensity of sewage outfalls and their effects on water quality and to find the relation between chemical and biological conditions both during the two seasons present, and between upstream and downstream locations from outfalls. Furthermore, it is compared the accuracy of the indexes: the BMWP-Col and the ABI, to assess the biological water quality in the rivers of the high Andes of Ecuador. The results may potentially be used as a base line for further research and scenario analysis for the restoration of the Cuenca River basin.

The outline of this chapter's approach is shown in Fig. 3.1 and clarifies its development in five different steps. The Introduction (section 3.1) discusses the first step of problem definition and the second step of planning and conceptual influence. The following step of data collection and processing is shown in sections 2.2.1 and 2.2.3. The data analysis is present in section 3.2, while step 4 is divided in 6 different approaches that are presented in the Results (section 3.3). Step 5 is addressed in 4 different ways that are analyzed in section 3.4, namely Discussion. The final step, Conclusion, is described in section 3.5.

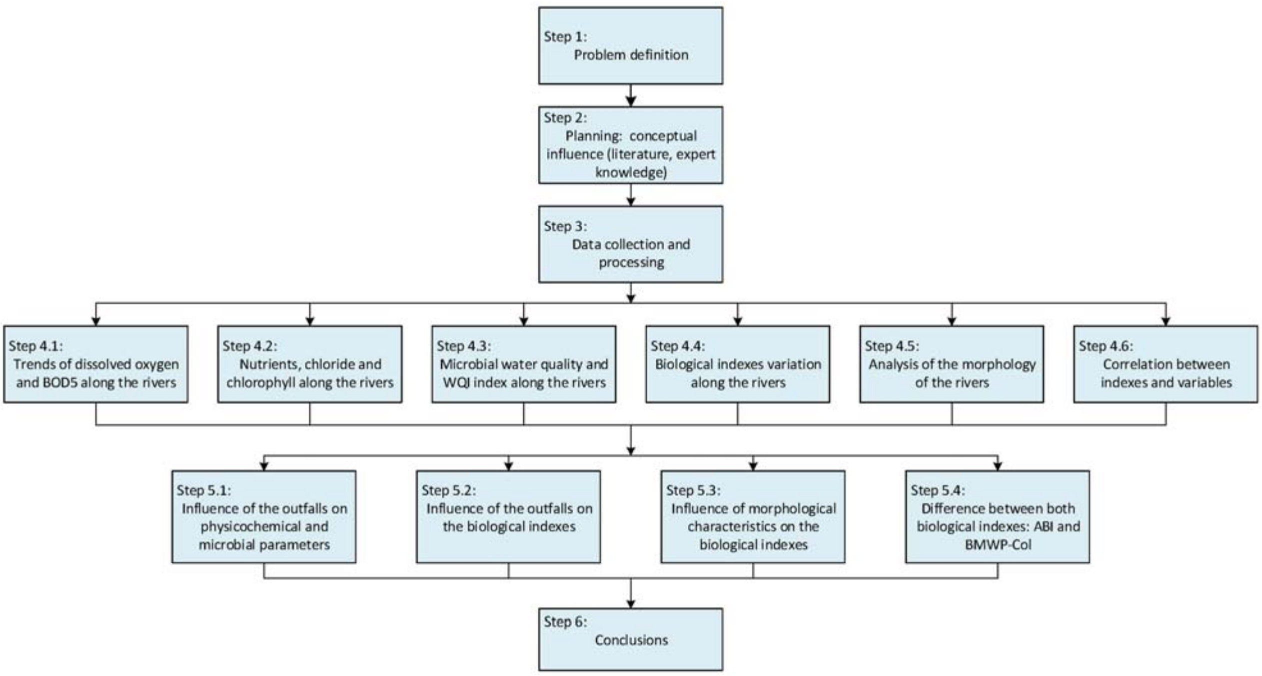


Fig. 3.1. An outline of the chapter's approach.

3.2 Data analysis

From the physicochemical variables measured, the main parameters that could have influenced the variation of the biological index in an urban area were analyzed, in addition to their performance along the four rivers in both rainy and dry seasons. The parameters considered were the five-day biological oxygen demand (BOD_5), ammonium (NH_4), nitrate (NO_3), nitrite (NO_2), orthophosphates (PO_4), chloride (Cl) and fecal coliforms (FC). Included were parameters that could be affected by pollution such as chlorophyll-a, dissolved oxygen (DO), oxygen saturation, temperature and the biological indexes (ABI and BMWP-Col). Because the Cuenca River originates as a result of the union of the Tomebamba and the Machangara Rivers, and in order to facilitate the description and the understanding of the results, the sites sampled in the Cuenca River are described as sites in the Tomebamba. The knowledge of the range of concentrations of these contaminants will be useful in a future analysis of possible scenarios to improve the water quality conditions of the streams that pass through the urban and suburban areas of Cuenca.

Additionally, the Water Quality Index (WQI) developed in 1970 by the US National Sanitation Foundation was calculated (Brown et al. 1970). The index is obtained on the basis of nine water quality parameters: oxygen saturation, fecal coliform, pH, BOD_5, temperature difference (over a distance of 1.6 km), total phosphate, nitrate, turbidity and total solids. Each parameter, according to its concentration receives a quality value that is multiplied by a weight given per each parameter. The sum of the weights of the nine parameters is 1 and the quality of each range from 0 to 100. The overall WQI is the sum of the weighted quality of each of the nine parameters. Water is graded for its physicochemical and microbiological quality in five categories from 0 to 100: 0-25 Very bad, 25-50 Bad, 50-70 Medium, 70-90 Good, and 90-100 Excellent.

To be able to compare the characteristics in each location, the hydro-morphological index of diversity (HMID) was calculated (Gostner et al. 2013). This index was developed to assess the hydro-morphological heterogeneity of alpine gravel-bed streams through the variation of hydraulic variables: flow velocity and depth in a cross section. The heterogeneity of a river site is greater when the HMID is higher. The HMID was selected because the rivers in the study area were mainly constituted by gravel-bed.

Finally, to assess the correlation between the ABI, the BMWP-Col and the WQI indexes and the 38 physicochemical parameters, the non-parametric Spearman rank correlation coefficient was computed. This coefficient is regularly used in ecology because it makes no assumptions about linearity between the two variables analyzed (Zuur et al. 2009, Nguyen et al. 2014). The independence of the data collected from the same sampling sites during both seasons was assessed based on the ABI and BMWP-Col values using a two-tailed Student *t*-test with a confidence level of 5%

(Witten and Frank 2005). By contrast, the dependence of the biological indexes between before and after outfalls was verified with a two-tailed paired Student *t*-test and with the aforementioned significance levels. Prior to the *t*-test analysis, the values of the biological indexes were verified to be normally distributed by means of the Shapiro-Wilks test (Thode 2002), as well as the homogeneity of their variance that was checked using the Levene test (Anderson 2006). The correlation analysis, the Student *t*-test and the scatterplots were performed in R (R-Core-Team 2017).

3.3 Results

3.3.1 Dissolved oxygen (DO), oxygen saturation, temperature and biological oxygen demand (BOD_5)

The DO concentration in the four main rivers was close to the saturation level, and its values were mostly higher in dry season compared to rainy season. However, when the trend of this parameter (Fig. 3.2A) was analyzed along the Tomebamba River, it was noted that in rainy season, after joining with the Yanuncay River, the DO concentration decreased to values about 7.5 $mg.L^{-1}$. However, during dry season, the variation of DO along the Tomebamba fell steadily from the highest sampling site (To18 – Fig. 2.1) until the junction with the Yanuncay, a point from which the concentration was around 8 $mg.L^{-1}$.

In the Machangara River, the DO concentration (Fig. B1A) remained almost constant in both seasons, with a minimum of 8 $mg.L^{-1}$. During both seasons, in the Yanuncay River, the DO declined after joining with the Tarqui River. However, in rainy season, an erratic fluctuation was registered when the Yanuncay crossed the city (Fig. B2A), which could have been due to the discharges from the CSOs as well as the surface water outfalls (SWO). In the Tarqui River, during both seasons, the DO concentration (Fig. B3A) was a little lower upstream than downstream. The lowest measurement of Dissolved Oxygen was registered in one of the tributaries of the Tarqui called the Salado Brook (Ta02 – Fig. 2.1), with a value of 2.2 $mg.L^{-1}$.

Oxygen saturation in the main four rivers was approximately 100% in both seasons, with higher levels upstream from Cuenca than downstream in the Tomebamba and in the Yanuncay (Fig. 3.2B and Fig. B2B). Yet in the Machangara, the oxygen saturation was almost stable along the river (Fig. B1B). Conversely, oxygen saturation in the Tarqui River experienced the opposite trend compared to the other rivers, with its lowest concentration upstream rather than downstream in both seasons (Fig. B3B). No differences in the tendency of oxygen saturation were found between both seasons, with the exception of the Machangara River, where the concentrations were higher in the rainy than in dry season.

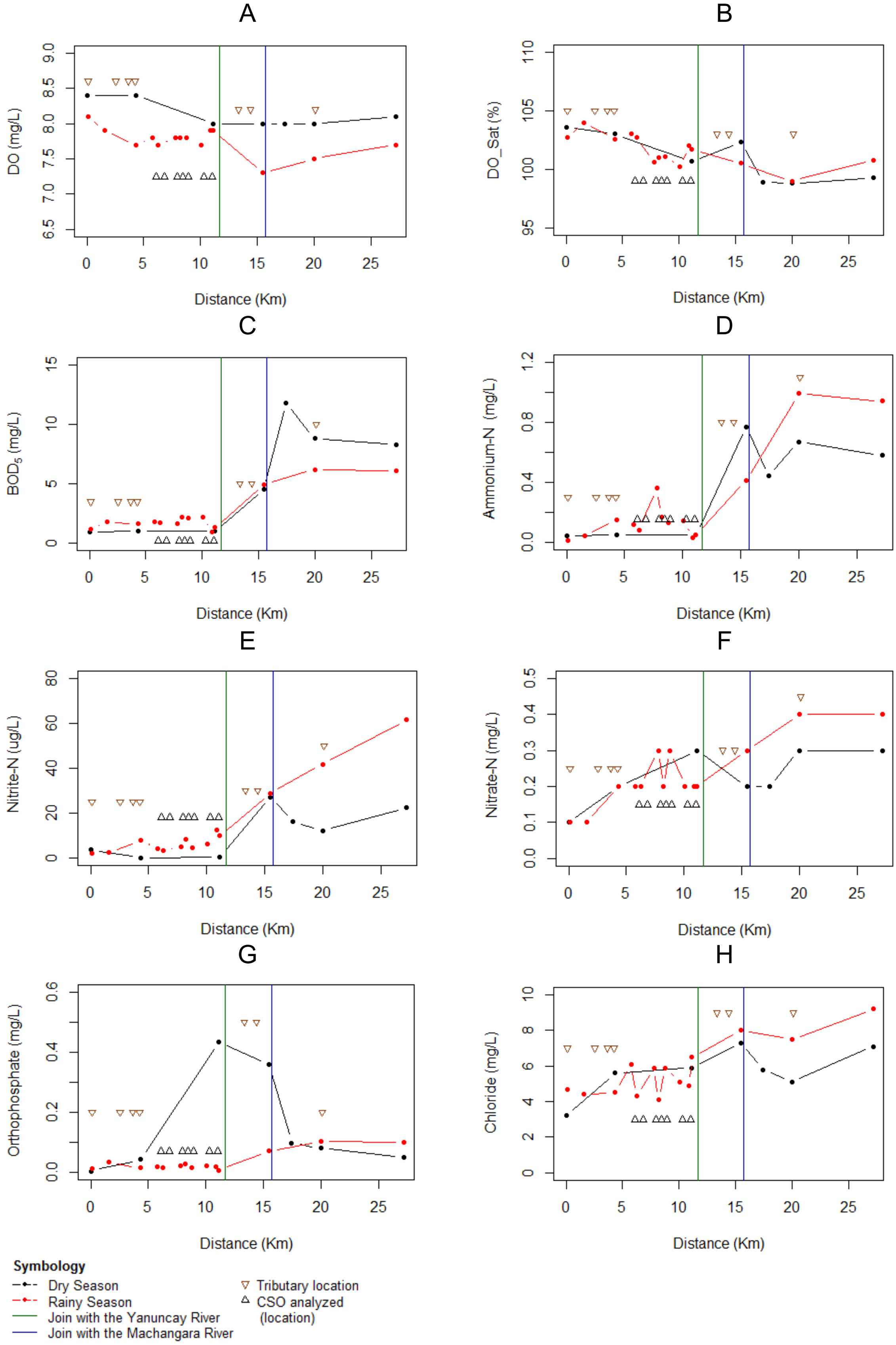


Fig. 3.2. Parameter variations along the Tomebamba River: (A) DO, (B) DO-Saturation, (C) BOD_5, (D) Ammonium, (E) Nitrite, (F) Nitrate, (G) Orthophosphate, (H) Chloride. The parameters are expressed in different scales.

The BOD_5 was higher in rainy season in the Tarqui River. During both seasons, from where the Tarqui flowed into the city onward (Ta23 – Fig. 2.1), the concentration of BOD_5 rose markedly (Fig. B3C), due to the discharges from the Salado Brook (Ta02 – Fig. 2.1), a highly polluted tributary.

The BOD_5 registered in stretches where the CSOs discharged, ranging from 1.2 to 2.2 $mg.L^{-1}$ and from 2.3 to 3.8 in the Tomebamba (Fig. 3.2C) and in the Yanuncay (Fig. B2C) Rivers, respectively.

3.3.2 Nutrients, chloride and algae

Ammonium (NH_4) in the Tomebamba River (Fig. 3.2D) had the following trend in both seasons: the concentration was close to zero in the highest sampling site of the study area, increasing slowly until the river met the city (To12 – Fig. 2.1). The ammonium value remained steady until the Tomebamba River joined the Yanuncay River, registering a peak value (To42 – Fig. 2.1) during rainy season that was probably due to the discharge of a CSO (ToCr17-Fig. 2.1). From the confluence with the Yanuncay, ammonium increased sharply in the Tomebamba until joining the Machangara River, which was due to its highly contaminated tributary, the Valle Brook. In dry season, after the confluence with the Machangara River, the ammonium fluctuated with values from 0.4 to 0.7 mg/L. While in rainy season, ammonium in the Tomebamba River increased rapidly until the discharge of the U-WWTP (Cu03 – Fig. 2.1); from there, the values declined slowly. Generally, the ammonium concentrations were higher in the rainy season than in the dry season.

In the Machangara River (Fig. B1D), ammonium concentrations were lower or equal to 0.2 mg/L in both seasons, with higher concentrations in dry season. During both seasons in the Yanuncay River, the ammonium (Fig. B2D) was below or equal to 0.1 mg/L until uniting with the Tarqui River. From this junction, the measured ammonium concentration was higher in rainy season. In both seasons, the ammonium in the Tarqui River (Fig. B3D) reached a concentration of 0.1 mg/L until the river arrived at the city (Ta23). However, from this point, ammonium in the Tarqui increased more sharply in dry season, due to the discharge of its polluted tributary, the Salado Brook.

In all four rivers, nitrite was the only pollutant whose concentration was equal to or higher in rainy than in dry season. During the dry season, nitrite in the Tomebamba River (Fig. 3.2E) had a value around zero in the city before joining with the Yanuncay River, with small peaks in rainy season (Fig. 3.2E); values that probably originated from the CSOs operation. From the confluence with the Yanuncay, during both seasons, the values of nitrite increased markedly until joining with the Machangara River. From this point onward during the dry season, the values in the Tomebamba declined until after the discharge of the U-WWTP (Cu03 – Fig. 2.1), at which point the values started to rise slowly. Conversely, during the rainy season, in the Tomebamba

River, the concentration of nitrite increased sharply after meeting the Machangara River (Cu02 – Fig. 2.1).

In the Machangara River, nitrite had the same pattern in both seasons (Fig. B1E), with higher values in rainy season. In the Yanuncay River during dry season, nitrite increased gradually (Fig. B2E) until prior to meeting the Tomebamba; while in rainy season, nitrite rose slowly until the Yanuncay arrived at the city of Cuenca. There, the concentration increased steadily to a value of 40.5 mg/L before joining the Tomebamba River. In the Tarqui River (Fig. B3E), nitrite had a similar trend during both seasons with higher concentrations in the rainy season. The values remained steady until the Tarqui reached Cuenca, from which point a marked increase of nitrite to a concentration of about 50 mg/L was noted in both seasons.

There was a steady increase of nitrate along the Tomebamba River (Fig. 3.2F) in the dry season until prior to joining the Yanuncay River. From this point onward, the concentration of nitrite decreased until meeting the Machangara River, where nitrate stabilized until the discharge of the U-WWTP. This led to an increase of nitrate to 0.3 mg/L in the last stretch of the Tomebamba River. During rainy season, two peaks of nitrate were measured when the Tomebamba passed through the city, which could have been due to the operation of a combined sewage overflow (CSO). Where the Tomebamba met the Yanuncay River, the values grew steadily until 0.4 mg/L after the discharge of the U-WWTP, at which point the concentration of nitrate stabilized.

Nitrate in the Machangara River (Fig. B1F) was stable in rainy season with values higher than in the dry season. In the Yanuncay River (Fig. B2F), during dry season, nitrate increased slowly until joining the Tomebamba. However, in rainy season, this pollutant in the Yanuncay fluctuated until before the junction with the Tarqui, maintaining its concentration from here until joining the Tomebamba. In the Tarqui River during dry season, nitrates (Fig. B3F) increased until the river arrived at the city, at which point the concentration declined slightly before meeting the Yanuncay. In rainy season, nitrate along the Tarqui had concentrations around 0.4 mg/L.

Along the Tomebamba River, orthophosphate (Fig. 3.2G) was almost zero in rainy season with a slight increase to 0.1 mg/L after joining with the Yanuncay River. In dry season, this pollutant was almost zero upstream from Cuenca, increasing its concentration to 0.4 mg/L in the urban area before joining with the Yanuncay, after which the value remained stable until meeting the Machangara River. At this point, the concentration dropped slowly until about zero.

Orthophosphate along the Machangara River registered at about zero during both seasons (Fig. B1G). Similar values were measured in the Yanuncay River (Fig. B2G) in both seasons, until meeting the Tarqui River, from which point onward the concentration increased to 0.1 mg/L before joining the Tomebamba. In the Tarqui River, higher orthophosphate concentrations were registered in the rainy season (Fig.

B3G). However, upstream from the city, the concentrations of orthophosphate were the same during both seasons.

Chloride had a similar trend in the Tomebamba River in both seasons (Fig. 3.2H), with a lower concentration upstream from Cuenca, registering a steady increase after the confluence with the Yanuncay River and a decrease after joining the Machangara River. Some peaks were also shown in the Tomebamba during rainy season (To37, To42 and To44), apparently were produced as a result of the CSOs. In the Machangara River, chloride was higher in rainy season with an unchanged concentration in dry season (Fig. B1H). In the Yanuncay, the concentration of chloride (Fig. B2H) remained stable with a slight variation in rainy season; where it registered a peak in Ya36 (Fig. 2.1) that could have been due to the operation of a CSO; there was an increase of this pollutant after the confluence with the Tarqui. The Tarqui River showed the highest concentration of chloride (Fig. B5H) during both seasons. In fact, this pollutant was higher in dry season upstream from the city. The two tributaries that transport wastewater: the Valle Brook in the Tomebamba and the Salado Brook in the Tarqui, had values higher than 50 $mg.L^{-1}$.

The algae presence was measured as chlorophyll-a only in dry season. The levels of this parameter in the Tomebamba River (Fig. 3.3A) were stable until the confluence with the Machangara River. Nevertheless, from this point, the concentration increased sharply to a value of 15.8 µg/L until the discharge by the U-WWTP. At that point, there was a slight fluctuation in the last analyzed stretch of the Tomebamba River. In the Machangara River, the chlorophyll-a (Fig. B5A) had values lower than 4 µg/L. The Yanuncay River contained concentrations of chlorophyll-a (Fig. B5B) that fluctuated from 12 to 16 µg/L. In the Tarqui River (Fig. B5C), the concentration of chlorophyll-a diminished from 13.5 µg/L in the highest studied sampling site upstream to 10 µg/L before joining with the Yanuncay.

3.3.3 Microbial water quality and the water quality index (WQI)

Fecal coliforms (FCs) were measured as most probable number per 100 mL ($MPN.100\ mL^{-1}$) and expressed in a logarithmic scale in the analysis of their variation in the rivers. In the Tomebamba River (Fig. 3.3B) during dry season, upstream from Cuenca, the FCs had the lowest concentration. Their values increased gradually until the Tomebamba arrived at the city, remaining steady from this point with a slight variation until joining the Yanuncay River. Downstream from the latter place, the FCs rose steadily in the Tomebamba to 5.4E+06 MPN (Cu04 – Fig. 2.1) until prior to the discharge of the U-WWTP, a point in which the FCs started to reduce steadily along the river. In rainy season, FCs had higher concentrations upstream from Cuenca and lower values than in dry season downstream from the confluence of the Tomebamba and the Yanuncay, with a peak of 3.3E+06 MPN after the discharge of the U-WWTP.

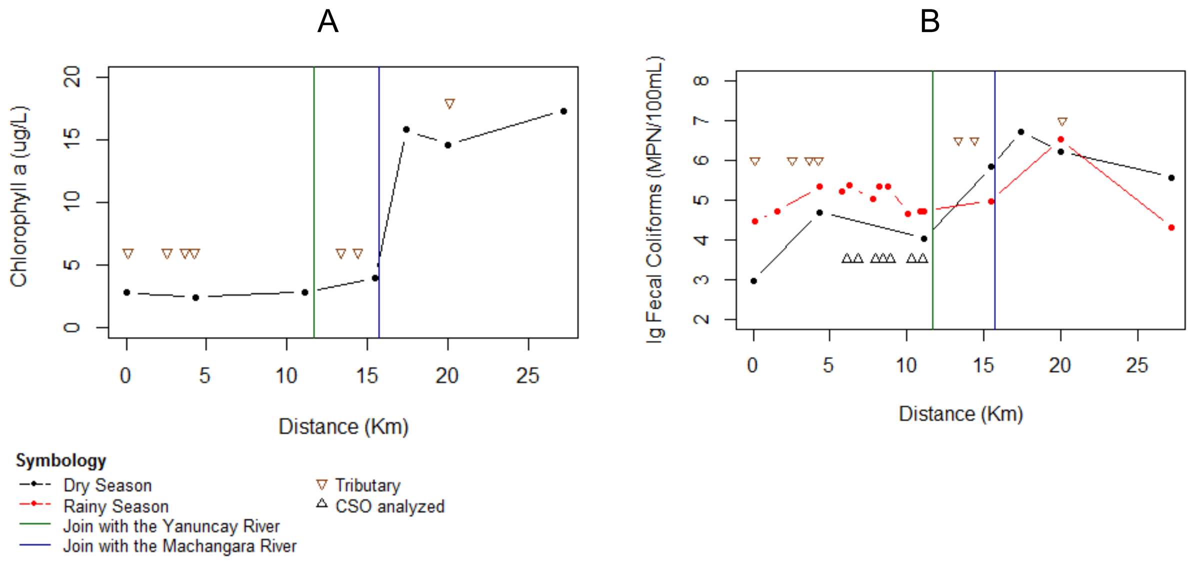


Fig. 3.3. (**A**) Chlorophyll-a and (**B**) fecal coliforms expressed in logarithmic scale, along the Tomebamba River.

In the Machangara River (Fig. B6A), during dry season, the FCs were lower upstream from Cuenca than in rainy season. When the river entered the city (Mc11 – Fig. 2.1) and flowed downstream (Mc13 – Fig. 2.1), the FCs increased continuously until 1.1E+0.6 MPN in dry season, values that were higher than in rainy season.

During both seasons, the FCs in the Yanuncay River (Fig. B6B) had similar trends, with lower concentrations upstream from Cuenca, and higher concentrations before the confluence with the Tomebamba River. During rainy season, by contrast to the other rivers, the concentrations of the FCs in the Yanuncay were higher than during dry season.

In the Tarqui River (Fig. B6C), the behavior and concentrations of fecal coliforms was similar in both seasons, with a higher concentration of FCs prior to joining the Yanuncay than upstream from the city.

According to the values of the water quality index (WQI) shown in Table 3.1, the Tomebamba River upstream from the confluence with the Yanuncay River, had values higher than 70 (good water quality) during both dry and rainy seasons, while downstream from this union, the WQI was lower than 70 (medium water quality), with values slightly higher in rainy season. Along the Machangara River, the WQI was higher than 70 during dry season, whereas in rainy season the WQI was around 70. The WQI along the Yanuncay River registered values higher or equal to 70 during dry season, although during rainy season the WQI was lower than 70 both upstream and downstream from the city, registering more than 70 along the city. Along the Tarqui River the WQI was around 70 during dry season, while in rainy season this index fell below than 70. Only in polluted tributaries such as the Valle and Salado Brooks (To15 and Ta02 - Fig. 2.1) the WQI scored lower than 50 (bad water quality).

3.3.4 Morphology

With regard to the land use, 73% of the sites were sampled in the urban area, 22% in an arable area or livestock land and 5% in a forest. Submerged macrophytes were frequently found in 5% of the sampling sites, such as Ya17 (Fig. 2.1). Whereas they were common in only 2% and occasionally seen in 9% of the samples.

In sampling sites To15 and To16 (Fig. 2.1), the sludge layer in the bottom of the rivers was between 5 to 20 cm, while in Cu01, Mc13 and Ta04 (Fig. 2.1), the sludge was less than 5 cm. In the rest of the sampling sites, the mineral substrate was mostly composed of boulder (47%), cobble (28%), and gravel, sand and silt (25%). The bed compaction of the studied stretches varied from packed – unarmored (13%), to moderate (34%) and low (45%). The sediment angularity in the sampled sites was mainly composed of rounded (45%), well-rounded (31%) and sub-angular (19%), as well as cobble and angular (5%). Thus, different bed compaction and dissimilar sediment angularity than either upstream or downstream locations was found in sampling sites To12, Ya16 and Ya35 (Fig. 2.1). The hydro-morphological index of diversity (HMID) varied from 1.8 to 5.8, with the highest values upstream from the city (Table 3.1).

During dry season, in sampling sites (Fig. 2.1) such as To13 and Mc11, the velocity was higher than their upstream locations of To19 and Mc09 respectively. Whereas, during rainy season, similar findings were found in sampling sites To44 and To31 that registered higher velocity than their downstream locations, To43 and To18, respectively. The aforementioned downstream sampling sites generally had a higher HMID than their upstream locations, with the exception of sites To13 and To31 (Table 3.1).

3.3.5 Biological indexes: Andean Biotic Index (ABI) and Biological Monitoring Working Party adapted to Colombia (BMWP-Col)

During both seasons, the values of the ABI and BMWP-Col in the four rivers: Tomebamba, Machangara, Yanuncay and Tarqui, upstream and downstream from Cuenca, as well as before and after the main directed outfalls and from the most important overflows in the Tomebamba and in the Yanuncay Rivers, were determined and shown in Table 3.1 and Fig. B7.

According to the Student-test, when the same sampling sites between both seasons were compared, the overall number of taxa and the Shannon-Wiener index, were shown to be independent ($p < 0.05$). However, the BMWP-Col appeared to be marginally independent, but significantly higher during rainy season; the ABI and the number of sensitive taxa (NST), excluding the Ephemeroptera-Plecoptera-Tricoptera-taxa, were marginally significantly higher in rainy season. The results of the paired Student *t*-test showed that the values of both biological indexes after an outfall were dependent on their values prior to this discharge. However, in dry season, the values of NST were shown to be independent both before and after an outfall.

In the Tomebamba River, upstream from Cuenca (To18 – Fig. 2.1), during the dry season the biological quality of the river measured with the ABI and BMWP-Col was good. However, this quality gradually decreased to moderate until the Tomebamba arrived at the city (To12 – Fig. 2.1). When this river passed through Cuenca, both biological indexes fell markedly to bad before joining with the Yanuncay (To19 – Fig. 2.1). In the next stretch of the Tomebamba, there was a slow rise of the ABI and BMWP-Col to deficient, a classification that was registered before joining the Machangara River (To13 – Fig. 2.1). This despite the polluted stream, the Valle Brook (To15 – Fig. 2.1), that discharged in this stretch. Downstream from the Machangara, the ABI and BMWP-Col in the Tomebamba decreased slightly to the category of bad and deficient respectively, before the discharge of the U-WWTP (Cu04 – Fig. 2.1), showing a slow recuperation of both biological indexes in the next stretch to a quality of moderate (Cu03 – Fig. 2.1), a classification that fell again in the last study tract of the Tomebamba River (Cu02 – Fig. 2.1).

During the rainy season, the Tomebamba River showed a different trend than in the dry season. Upstream from Cuenca, the biological indexes were similar in dry season, classifying in the range of good biological conditions. Yet, when the Tomebamba passed through Cuenca and upstream from that point, the values of the ABI and BMWP-Col and their classifications, were higher than those registered in dry season (Fig. B7). These biological indexes fluctuated along the river, with a tendency to reduce their values steadily downstream. However, as can be seen in Table 3.1, at some sampling sites with different morphological characteristics, the ABI and BMWP-Col were higher than previous points upstream (To31, To12, To44 and To46).

In the Machangara River, during dry season, the ABI and BMWP-Col were higher upstream from Cuenca than in rainy season (Mc09 and Mc11 – Fig. 2.1); while upstream from the city, these biological indexes had lower values in dry season (Table 3.1 and Fig. B7). An interesting pattern was found during both seasons: in the highest analyzed sampling site in the Machangara River (Mc09), the ABI and BMWP-Col had lower values than the downstream site (Mc11). Yet, from Mc11 to downstream from the city, the biological indexes' tendencies were different in both seasons, showing a constant reduction of their values in dry season while in rainy season, the ABI and BMWP-Col increased gradually. The qualification of the biological indexes along the Machangara River in dry season varied from moderate to good and from good to deficient, yet in rainy season, this condition varied from deficient to moderate.

Table 3.1. Biological water quality based on the ABI and BMWP-Col invertebrate community indexes, physicochemical water quality base on the WQI index, and the hydro-morphological qualification based on the HDIM index, both upstream and downstream from the main outfalls in the four main rivers of the Cuenca River basin (information two seasons).

Outfall name	Upstream					Downstream					Biological Index difference	
	sampling sites	ABI	BMWP -Col	HMID	WQI	sampling sites	ABI	BMWP -Col	HMID	WQI	ABI	BMWP-Col
Tomebamba River												
Dry season												
Scattered population	**TO18**	76	85	5.5	82	TO12	47	50	1.9	80	29	35
Discharges from some bad regulated overflows in the city	**TO12**	47	50	1.9	80	TO19	10	12	3.3	76	37	38
Confluence with Yanuncay River, TO15 and TO16	**TO19**	10	12	3.3	76	TO13	20	34	2.9	70	-10	-22
Confluence with Machangara River and scattered population	**TO13**	20	34	2.9	70	CU04	14	26	3.0	62	6	8
WWTP-Ucubamba	CU04	14	26	3.0	62	CU03	39	46	3.8	66	-25	-20
Scattered population	**CU03**	39	46	3.8	66	CU02	13	20	3.8	58	26	26
Rainy season												
Scattered population	**TO18**	61	64	3.1	73	TO31	103	113	2.3	73	-42	-49
TO32 and TO34	TO31	103	113	2.3	73	TO12	122	143	2.8	72	-19	-30
TO35 and overflow TOCL08	**TO12**	122	143	2.8	72	TO37	88	90	2.3	72	34	53
Overflow TOCR12	TO37	88	90	2.3	72	TO38	84	102	4.3	71	4	-12
Overflow TOCR17	TO38	84	102	4.3	71	TO42	51	59	4.4	75	33	43
Overflow TOCR21	TO42	51	59	4.4	75	TO43	54	60	4.4	73	-3	-1
Overflow TOCR24	TO43	54	60	4.4	73	TO44	80	85	4.7	71	-26	-25
Overflow TOCL25	TO44	80	85	4.7	71	TO45	37	47	3.9	71	43	38
Overflow TOCL30	TO45	37	47	3.9	71	TO46	65	80	2.8	74	-28	-33
Overflow TOCL32	TO46	65	80	2.8	74	TO19	55	66	2.8	71	10	14
Confluence with Yanuncay River, TO15 and TO16	**TO19**	55	66	2.8	71	TO13	38	57	2.7	67	17	9
Confluence with the Machangara River, and discharge from U-WWTP	**TO13**	38	57	2.7	67	CU03	37	49	2.4	71	1	8
Scattered population	**CU03**	37	49	2.4	71	CU02	33	48	4.5	67	4	1
Machangara River												
Dry season												
Agricultural runoff	**MC09**	41	50	2.3	74	MC11	62	68	4.2	79	-21	-18
Some industrial discharges	**MC11**	62	68	4.2	79	MC13	23	21	3.6	74	39	47
Rainy season												
Agricultural runoff	**MC09**	18	24	3.8	71	MC11	35	40	4.1	69	-17	-16
Some industrial discharges	**MC11**	35	40	4.1	69	MC13	49	55	4.6	76	-14	-15
Yanuncay River												
Dry season												
Agricultural runoff	**YA15**	33	41	5.8	75	YA16	68	68	3.1	73	-35	-27
Confluence with Tarqui River	**YA16**	68	68	3.1	73	YA17	35	40	3.1	70	33	28
Rainy season												
Agricultural runoff	**YA15**	82	72	4.7	66	YA30	59	58	1.9	76	23	14
Overflow YACR03	YA30	59	58	1.9	76	YA34	16	20	2.4	73	43	38
Overflow YACR08	YA34	16	20	2.4	73	YA35	56	76	3.9	73	-40	-56
Overflow YACL13	YA35	56	76	3.9	73	YA36	25	41	2.7	72	31	35
Overflow YACR13	YA36	25	41	2.7	72	YA37	53	63	2.7	72	-28	-22
Overflow YACR17	YA37	53	63	2.7	72	YA16	29	44	1.8	75	24	19
Confluence with Tarqui River	**YA16**	29	44	1.8	75	YA17	52	70	3.0	66	-23	-26
Tarqui River												
Dry season												
Agricultural runoff	**TA05**	30	48	3.5	72	TA23	22	39	2.8	73	8	9
TA02 and TA23	**TA23**	22	39	2.8	73	TA01	17	32	4.2	68	5	7
Rainy season												
Agricultural runoff	**TA05**	33	50	3.1	63	TA23	28	47	2.4	64	5	3
TA02 and TA23	**TA23**	28	47	2.4	64	TA01	23	34	2.6	58	5	13

Bolded sites were sampled in both seasons.
ABI = Andean Biotic Index; BMWP-Col = Biological monitoring working party adapted to Colombia; HIMD = hydro-morphological index of diversity; WQI = Water quality index developed by the US National Sanitation Foundation.

In the Yanuncay River, during dry season, the biological Indexes had similar values upstream (Ya15 – Fig. 2.1) and downstream from Cuenca (Ya17 – Fig. 2.1).

Conversely, during rainy season, both indexes showed different patterns, thus the ABI was higher upstream and the BMWP- Col was similar in both sampling sites (Fig. B7). However, in dry season, Ya16 (Fig. 2.1) with its location before the Tarqui River, had a different morphology and had higher biological index values than upstream and downstream sampling sites (Table 3.1). The qualification of the ABI and BMWP-Col in dry season, remained good until the Yanuncay joined with the Tarqui, a point where this quality changed to deficient for the ABI and to moderate for the BMWP-Col. Similarly, in rainy season, when the Yanuncay passed through Cuenca, in some sampling sites with different morphological characteristics, both biological indexes were higher than their previous points (YA35, YA37 and YA17 - Table 3.1). The qualification of the ABI and BMWP-Col in rainy season, varied from good quality in the highest analyzed sampling site upstream from Cuenca, to moderate and from moderate to deficient and vice versa, downstream along different stretches of the Yanuncay.

In the Tarqui River, the biological indexes had similar trends during the seasons with a slightly higher value in rainy season (Table 3.1 and Fig. B7). Upstream from Cuenca, the biological water quality was better than downstream from the city. During both seasons, the BMWP-Col index was in the range of moderate and deficient, while according to the ABI, the biological water quality remained deficient in both seasons.

Generally, the biological water quality ranged higher with the BMWP-Col than the ABI. The difference of these indexes was higher in the Tarqui River than the other rivers in both seasons. However, there were some sampling sites, four during dry season and five during rainy season that showed different patterns. During dry season, one sampling site in each of the following rivers: Tomebamba (To15), Machangara (Mc13), Yanuncay (Ya16) and Tarqui (Ta03), revealed that the ABI was higher or equal to the BMWP-Col. While, in rainy season, the Tomebamba and the Yanuncay had three (To17, To32 and To35) and two (Ya15 and Ya3) sampling sites with higher values of the ABI, respectively. The absolute difference between the biological indexes from upstream and downstream of an outfall was 20.8 for the ABI and 22.5 for the BMWP-Col, while their standard deviations were 13.8 for the ABI and 15.5 for the BMWP-Col.

The higher ABI and BMWP-Col registered in the majority of the sampled sites during rainy season, was mainly governed by a higher number of taxa in those sites. This, in addition to a higher quantity of Ephemeroptera-Plecoptera-Tricoptera (EPT) families and other sensitive taxa that were found. The presence of *Hydroptilidae*, *Calamoceratidae* and *Hydrobioscidae* were higher in rainy season, increasing from four to 16, from one to 12, and from nine to 16 in sampling sites, respectively. Likewise, *Hyallelidae*, *Simuliidae*, *Scirtidae* and *Elmidae* increased their presence from 10 to 21, from 13 to 31, from three to 13, and from 11 to 23 sites, respectively. *Perlidae* was present in four sampling sites during rainy season, while in dry season, this taxon was not found. However, the presence of *Baetidae* increased in rainy season from 25 to 29 in sampling sites.

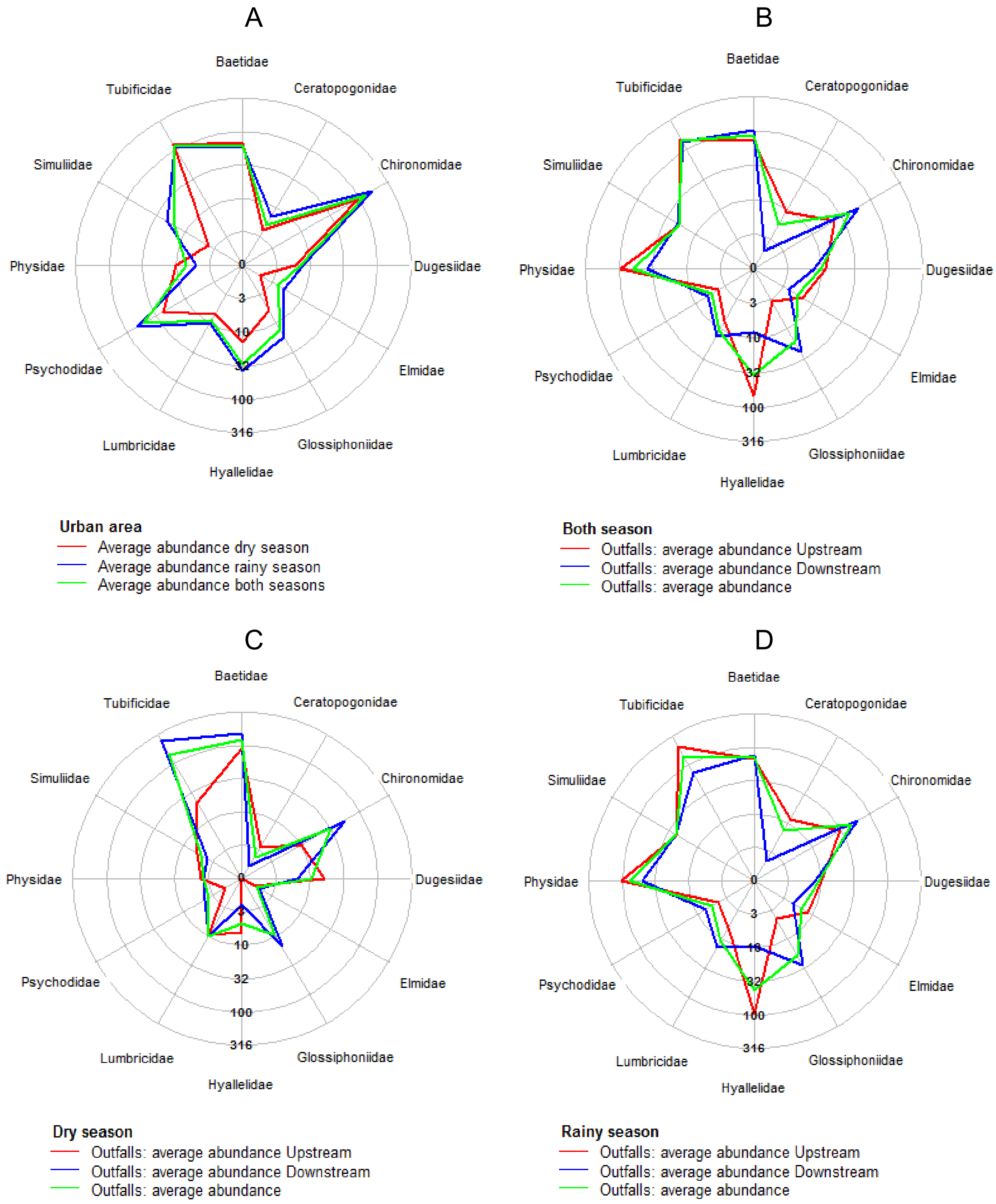


Fig. 3.4. Average abundance in log scale: (**A**) Urban area in both seasons, (**B**) Upstream and downstream outfall in both seasons, (**C**) Upstream and downstream outfall during rainy season (**D**) Upstream and downstream outfall during dry season.

With regard to the taxa abundance between seasons (Fig. 3.4A - Table B1 and Table B2), *Chironomidae*, *Tubificidae*, *Baetidae* and *Physidae* had the highest prevalence in the sampling sites. Of these, *Tubificidae* had the highest abundance in rainy season, while *Chironomidae* was highest during dry season. Nevertheless, for

Elmidae, *Hyallelidae*, *Psychodidae* and *Simuliidae*, the abundance increased sharply from dry season to rainy season.

The abundance of *Ceratopogonidae*, *Hyallelidae* and *Physidae* between upstream and downstream from outfalls (Fig. 3.4B - Table B1 and Table B2) plummeted, while the abundance of *Glossiphoniidae* and *Psychodidae* rose. During dry season (Fig. 3.4C - Table B1 and Table B2), the abundance of *Tubificidae* and *Physidae* was lower upstream than downstream from an outfall, whereas in rainy season (Fig. 3.4D - Table B1 and Table B2) the opposite was true for these taxa.

3.3.6 Correlation analysis between ABI, BMWP-Col, WQI, chemical and microbiological parameters

In the next paragraph, it is described the main results from the correlation done with the non-parametric Spearman analysis between ABI and BMWP-Col with 11 variables, DO, DO saturation, BOD_5, Log-fecal-coliforms, orthophosphates, ammonium, nitrate, nitrite, chloride, water temperature, and chlorophyll-a, and the measurement obtained between the aforementioned 11 variables.

From these results, the ABI and BMWP-Col had absolute correlation values higher than 0.4, with oxygen saturation, BOD_5, orthophosphate, chloride, ammonium and nitrite. The correlation values between the physicochemical variable with the ABI were generally higher than the values that were obtained with the BMWP-Col, with the exception of ammonium. The BOD_5 exhibited higher values of correlation than 0.7 with nitrite, oxygen saturation and log-fecal coliforms. Nitrite had a Spearman correlation coefficient higher than 0.7 with nitrate, chloride and BOD_5. Important correlations higher or equal to 0.7 were also noticed between: WQI and BOD_5, DO and temperature, DO and oxygen saturation and orthophosphate and chloride. While, the WQI had a higher value with the ABI (0.45) than the BMWP (0.42), these values were much higher during dry season than in rainy season for both biological indexes. Thus, in dry season the correlation value between the WQI and the ABI was 0.64, while this value was 0.59 between WQI and BMWP-Col. Similarly, during dry season the WQI had a higher correlation with the ABI (0.40) than with BMWP-Col (0.35).

3.4 Discussion

3.4.1 Influence of the outfalls on the physicochemical parameters, chlorophyll-a and fecal coliforms

A characteristic observed in the analyzed rivers was that the upstream sites had lower concentrations of BOD_5 and pollutants than downstream sites, while the oxygen saturation was higher upstream than downstream. The concentrations of dissolved oxygen (DO), especially in the studied rivers, were close to the saturation level, a finding that was consistent with that described by Jacobsen (1998) and Jerves-Cobo et al. (2017). These authors mentioned that mountain rivers with their high velocity and turbulence can maintain a relatively high oxygen saturation level. On the other hand, the DO was higher in dry season than in rainy season. A similar trend in the water temperature was found by Ríos-Touma et al. (2011) in the high Andean altitude in northern Ecuador. In this regard, it is known that when water temperature increases, the solubility of oxygen decreases at the same partial pressure (Wilcock et al. 1978) leading to reduced DO concentrations. However, the daily temperature range registered in some studies in the Ecuadorian Andean rivers was around 5 ^{o}C (Jacobsen 2008), which was higher with monthly variations registered in this study (~1 ^{o}C). Furthermore, it was found that oxygen saturation did not have as strong a correlation with temperature as dissolved oxygen.

During the dry season in the study area, the BOD_5, nutrients and chloride increased along the rivers, from two possible point sources: either sewage discharges or incorrect calibration of operational levels of sewage overflows. However, in rainy season, in addition to the aforementioned sources, point overflow discharges and disperse pollution transported by surface runoff contributed to the pollution of the rivers. In the rivers studied, it was possible to identify that a sampling site with a lower DO had a higher concentration of either BOD_5, nitrite, nitrate, chloride, or orthophosphate. As a result, the depletion of oxygen is produced in rivers during aerobic conversion of organic wastes by bacterial decomposition (Rauch et al. 1998), which is one of the main effects of organic pollution: the reduction of the oxygen level in rivers (Hynes 1960). A similar trend was noted along the rivers where a higher correlation was found between oxygen saturation and the BOD_5 than with nutrients and chloride. These results agreed with the absolute values of the Spearman coefficient obtained from the correlation analysis. Additionally, the reduction of BOD in rivers happens in two ways: as the result of immediate oxygen consumption and with delayed oxygen consumption from the sedimented BOD at the bottom of the streams (Harremoës 1982). Regarding combined sewer overflows (CSOs), two effects on the DO concentration in rivers are known: the first is an immediate effect that is produced by the degradation in the water phase of an extended organic waste discharged, and the second, is a delayed effect formed by adsorption of the organic matter fixed to the bed of the river that has a long affection period - longer than an overflow discharge - lasting 24 hours (Harremoës 1982, Hvitved-Jacobsen 1982). In most of the sampling sites in this study, during the rainy season, the delayed dissolved oxygen effect was registered from the samples that were taken either when a rainfall event diminished its intensity, or when it stopped.

Despite the fact that nitrite was measured in a very low range and expressed as μg N/L, this pollutant showed a higher concentration during rainy season in all rivers. In this regard, Najafpour et al. (2008) suggested that the concentrations of this pollutant were found to be related to higher flows in the respective rivers. However, the nitrite accumulation is not the norm and its concentration depends on the diverse parameters that influence the balance of its formation and transformation (Philips et al. 2002).

Chloride is a relatively conservative substance and its principal sources in rivers are point discharges either from sewage systems, WWTP or CSOs, but this pollutant is less frequent as a result of diffuse catchment drainage (Brooker and Johnson 1984). In this chapter, a similar trend was found when polluted streams joined with a main river such as the Salado Brook with the Tarqui River, or when the CSOs discharged into the Tomebamba and Yanuncay Rivers, showing some chloride peaks in this process. However, the Tarqui River also brought chloride from upstream, probably due to the daily salt (NaCl) intake of cattle (McDowell 1996) in the region's main livestock area. Chloride in the Tarqui registered a lower concentration in rainy season, which may have been caused by dilution.

Chlorophyll-a, which represents the measurement of an algae presence along the rivers, was more closely related to higher concentrations of BOD_5 and nitrite; features that were consistent with the results of the correlation analysis. A similar finding was described by Mallin et al. (2006) in three urban rivers and one creek, in a study done in 17 rivers in the state of North Carolina in the United States. These authors, also mentioned that the increase of chlorophyll-a and BOD could have been stimulated by nitrogen inputs. In this regard, in the last stretches of the Tomebamba River, where the concentration of chlorophyll-a was the highest, nitrogen could have been introduced either from isolated sewage systems or from U-WWTP discharges.

The fecal coliform (FC) concentrations in rivers are related to both upstream values and the addition of animal or domestic feces in their path, registering no other factors as significant (White and Dracup 1977). Mainly, peak increases are reflected during rainstorms, in which FCs are transported to the streams by runoff (White and Dracup 1977, Mallin et al. 2006, Arnone and Walling 2007). These features were seen in all rivers during both seasons. Thus in rainy season, the peaks from the CSOs were registered in both the Tomebamba and Yanuncay Rivers, as well as in some stretches of the rivers such as Cu01 in the Tomebamba. While in dry season, the FCs were influenced by badly regulated CSOs and some isolated sewage discharges. The FCs in the Tarqui River, which come mainly from the livestock area in the region, did not register an increase in concentration during the rainy season. This similar level of fecal coliform concentration found during the dry season was probably due grassland irrigation to being carried into the Tarqui River. The maximum concentrations of FCs were obtained during both seasons in polluted tributaries where a slow velocity (~0.3 m/s) was measured, in sites like the Milchichig and the Valle Brooks. In this regard, Wilkinson et al. (1995) indicated that areas in rivers with low velocities are ideal for the accumulation of fecal coliforms due to the presence of abundant macrophytes and sediments. However, in the river sampling sites studied, there was not an abundance of macrophytes. The same authors pointed out that a similar effect, with a lower accumulation of FCs is registered in the bare bed of rivers with slow velocity.

3.4.2 Influence of the outfalls in the biological indexes: the Andean Biotic Index (ABI) and the Biological Monitoring Working Party adapted to Colombia (BMWP-Col)

The ABI and BMWP-Col demonstrated a similar trend to that of oxygen saturation, which is a key factor in aquatic ecology (Jacobsen and Marín 2008); namely, points upstream from Cuenca that had a better biological water quality than downstream points. These results are consistent with the description of Hynes (1960), who mentioned that macroinvertebrate communities in rivers are affected by oxygen levels, alteration of the substratum and loss of available food sources, due to toxic pollutants such as ammonium and chloride. Additionally, because of low oxygen pressure and relatively high water temperatures found in Andean rivers, their macroinvertebrates are more sensitive to pollution than in temperate lowland streams (Jacobsen 1998). However, in some more polluted locations the ABI and the BMWP-Col were higher, which apparently correlated more to the oxygen saturation and dissolved oxygen than to the BOD_5 and pollutants concentrations. This hypothesis was confirmed with the Spearman correlation analysis and was consistent with the findings of Jacobsen and Marín (2008) and Friberg et al. (2010). Thus, the driving parameter that seemingly affected the values of both biological indexes was the oxygen saturation rather than the BOD_5 or nutrient concentrations, as was the case in the dry season between sampling sites To19 and To13, and in the rainy season between sites To44 and To43. In these downstream sampling sites, oxygen saturation did not drop similarly to other locations, possibly due to their higher velocity than in upstream sampling sites, allowing both a higher oxygen saturation, and a flushing out of the organic material stocked in the substrate of the stretch. The parameters that had higher influences in the ABI as well as the BMWP-Col were BOD_5, ammonium, orthophosphates and chloride, of which the first two factors were also found to be key variables for macroinvertebrates as noted by Friberg et al. (2010).

The ABI and the BMWP-Col were generally higher in rainy season in the same sampling sites in the urban area as well as downstream from these points, a finding that agreed with Burneo and Gunkel (2003) in the northern Andean region of Ecuador. This difference was given by the higher presence of sensitive taxa such as *Perlidae* and *Hyallelidae*. However, it was found that the improvement of both biological indexes during rainy season registered a similar value of oxygen saturation and either lower or higher BOD_5, nutrients and chloride concentrations. In one case, a lower concentration of these parameters was registered likely due to their greater dilution that diminished the effects of the pollution in the rivers. In the second case, a higher concentration of BOD_5, nutrients and chloride registered in certain sampling sites during the rainy season that may have been due to either point discharges from CSOs or runoff from the livestock area. In those sampling sites where a greater velocity was measured, it generated a washing out of the organic material accumulated in the substrate, such as in the urban area of the Tomebamba River, thereby improving the environmental conditions for macroinvertebrates. Similar outcomes were found by Jacobsen (1998) in the northern Andean region in Ecuador.

With regard to combined sewer overflows (CSOs) that operate during rainy events, the values of the ABI and the BMWP-Col were generally lower downstream than upstream from a CSO in the analyzed rivers of the Tomebamba and Yanuncay. These values were influenced especially by delayed oxygen depletion that affected the macroinvertebrate communities, an effect that was described in section 3.4.1. Moreover, lower concentrations of oxygen produced by this delayed effect in the river bottom were measured when the water surface registered almost saturated oxygen levels (Krejci et al. 1994). The oxygen levels that were registered in the water surface of the aforementioned stretches were almost saturated, and in connection with the taxa collected, the level of oxygen in the bottom of these river sections was likely very low. In fact, low concentrations could agree with a minor presence of sensitive taxa such as *Hyallelidae* and a higher abundance of less sensitive taxa such as *Chironomidae*. Yet, in two sampling sites: To46 (Tomebamba River) and Ya37 (Yanuncay River), after the CSOs discharged into ToCl30 and YaCr13, respectively (Fig. 2.1), the ABI and the BMWP-Col were higher than either upstream or downstream locations. These sites showed lower concentrations of BOD_5 and pollutants than surrounding points with similar levels of DO and oxygen saturation and comparable velocities. These lower concentrations were likely due to the CSOs that had lower flows with less pollution load, producing a softened impact on the rivers, and allowing for a faster recuperation of the aquatic environment.

The results shown in this chapter reflect the current condition in the lowland area of the Cuenca River basin. This information can be used by stakeholders to determine actions that could improve the current conditions. The implementation of possible actions to restore the water quality requires a scenario analysis that could give a quantitative prediction of its impact on river restoration (Mouton et al. 2009b). In future research, the analysis and optimization of different measures to improve the biological water quality could be conducted through an integrated ecological modelling, in which hydro-morphological pressure and physicochemical pollution can be assessed simultaneously (Holguin-Gonzalez et al. 2014).

3.4.3 Influence of morphological aspects in the biological indexes: ABI and BMWP-Col

The sampling sites in this study were mostly located in areas with human activities such as agricultural, livestock and urban land use, to evaluate the anthropic impact and overall, urban influence and outfalls effects in the rivers during dry and rainy seasons; features that agreed with the recommendations given by Strobl and Robillard (2008). Morphological characteristics in the rivers play an important role in the habitat structure of aquatic life (Burneo and Gunkel, 2003). For example, variability in the presence of macroinvertebrates has been found between sites with stones and vegetation versus stones alone and in areas with or without aquatic or marginal vegetation (Dallas and Day 2007). In mountain rivers, patterns that arise in more homogeneous rivers can be strongly masked by characteristics such as habitat availability, temperature and flow (Hawkins et al. 1997). In this context, during rainy season sampling sites such as Ya35 in the Yanuncay River and To12 in the Tomebamba River had higher values of both the ABI and the BMWP-Col than either upstream or downstream locations, with higher or similar concentration of BOD_5 and pollutants, with comparable levels of DO and oxygen saturation and with similar velocities. However, some sites such as To12 in the Tomebamba River and Ya35 in the Yanuncay River showed different morphological characteristics than the surrounding sampling sites. Thus, Ya35 and To12 had different bed compaction, dissimilar sediment angularity and higher HDIM than either upstream or downstream locations. In this regard, Holomuzki and Biggs (2003) mentioned that in high-gradient mountain rivers, sediment angularity and irregular shapes of stones diminish hydraulic strengths, offering micro-refuge to macroinvertebrates during high flow events. Yet, in Ya17 submerged macrophytes were frequent, providing an important habitat for many taxa (Dallas and Day 2007), characteristics that were not shown in neighborhood locations. Furthermore, the Ya17 had an HDIM higher than the upstream location. Apparently, morphological aspects had some influence on the score of the ABI and BMWP-Col in specific sampling sites such as To12, Ya17 and Ya35. These results were similar to those obtained by Jacobsen (1998), who reflected that a better performance relating to water quality and biological assessment was obtained in the dry season, in which morphological variables could play a more important role than pollutants. Moreover, these findings were confirmed with a much higher correlation between biological indexes and the WQI during dry season. However, the impact of CSOs can only be evaluated during rainy season.

3.4.4 The difference between both biological indexes: ABI and BMWP-Col

Both of the biological indexes used in this research, the ABI and the BMWP-Col, showed similar patterns along the rivers, generally the BMWP-Col scored higher than on the ABI. This was due to the fact that both indexes have different tolerance scores (TS) for various taxa. Thus, very sensitive taxa for BMWP-Col such as *Lymnaeidae*, *Sphaeriidae*, *Baetidae, Pyralidae* and *Psephenidae* (Álvarez 2005), ranged from low to medium sensitivity for the ABI (Encalada et al. 2011, Ríos-Touma et al. 2014). In this research, *Psephenidae* was present only during rainy season in a few points, while *Lymnaeidae* and *Sphaeriidae* were present during both seasons in sampling sites with low or moderate pollution. *Baetidae* also occurred frequently in both seasons, and this taxon was absent only in severely polluted points; a similar finding was mentioned by Jacobsen (1998) in his research in the northern Andes of Ecuador. For this reason, the TS given by the ABI for *Baetidae* could be more suitable than that given by the BMWP-Col for the high Andes region. *Pyralidae* was present in only one sampling site with moderate pollution in dry season. However, only *Helicopsychidae*, a very sensitive taxon, has a higher TS in the ABI index than the BMWP-Col (Álvarez 2005, Ríos-Touma et al. 2014). *Helicopsychidae* was found in one sampling site during rainy season in a point with very low pollution.

However, some taxa do not have a TS value, either for ABI, BMWP-Col or for both indexes, *Dugesiidae* and *Glossiphoniidae* do not have a TS for the ABI index (Ríos-Touma et al. 2014), taxa with a median sensitivity according to the BMWP-Col (Álvarez 2005). Unlike the BMWP-Col taxa such as *Acari* and *Dytiscidae* (Álvarez 2005) which were not scored, in the ABI index they were given a median TS (Encalada et al. 2011). *Lumbricidae* and *Stenopsychidae* have not been scored in either index (Álvarez 2005, Ríos-Touma et al. 2014). *Lumbricidae* was present in both seasons in sampling sites with moderate pollution, while *Stenopsychidae* was present in only one point with low pollution during rainy season. In a future update of the ABI index could be improved by adding the TS given by Acosta et al. (2009) of six and five for *Dugesiidae* and *Glossiphoniidae,* respectively. Similarly, this index could include TS to *Lumbricidae* and *Stenopsychidae*.

After analyzing both indexes, the ABI has proven to be more suitable for the Andes in Ecuador above 2400 m a.s.l, as a bioindicator of water quality. This could be attributed to the higher correlation values obtained between the WQI and the ABI than the BMWP-Col during both rainy and dry seasons. Moreover, the qualification given by WQI could be more suitable with biological water quality given by the ABI than the BMWP-Col, in sampling sites such as To13, Ya15 and Ta05.

3.5 Conclusions

The ecological water quality was assessed in both the dry and rainy seasons in the lowland area of the Cuenca River basin, where the city of Cuenca is located that receives high impact from human activities. During both seasons, the biological water quality measured with the Andean Biotic Index (ABI) and the Biological Monitoring Working Party adapted to Colombia (BMWP-Col) was better upstream than downstream from the city, agreeing with the concentration of the oxygen saturation and an inverse trend from BOD_5, nutrient and chloride concentrations. The biological quality was related more to the oxygen concentration and BOD_5 than nutrients. Both biological indexes had a clearer relationship with the BOD_5 and nutrient concentrations during dry season than in rainy season. In general, during the rainy season, the biological water quality was better, due to the presence of more sensitive taxa. In some locations, especially during rainy season, morphological variables such as sediment angularity, macrophytes and flow velocity were more influential on biological water quality than the concentration of pollutants. Nitrite was the only pollutant that had a higher concentration during rainy season in all rivers. Both biological indexes showed similar patterns along the rivers, generally the BMWP-Col scored higher than the ABI. The latter index seems to be more suitable for the high Andes region than the BMWP-Col as an indicator of the water quality. Yet, the ABI scoring could be improved to include new taxa. These results could now be used in a scenario analysis to improve the Cuenca River basin management in future studies.

Chapter 4. A Methodology to Model Environmental Preferences of Ephemeroptera—Plecoptera—Trichoptera (EPT) Taxa in the Machangara River basin (Ecuador)

Adapted from:

Rubén Jerves-Cobo, Gert Everaert, Xavier Iñiguez-Vela, Gonzalo Córdova-Vela, Catalina Díaz-Granda, Felipe Cisneros, Ingmar Nopens and Peter L. M. Goethals. (2017). A Methodology to Model Environmental Preferences of EPT Taxa in the Machangara River Basin (Ecuador). Water 2017, vol. 9, p. 195; doi:10.3390/w9030195. ISSN 2073-4441.

Abstract

Rivers have been frequently assessed based on the presence of the Ephemeroptera—Plecoptera—Trichoptera (EPT) taxa in order to determine the water quality status and develop conservation programs. In this chapter is evaluated the abiotic preferences of three families of the EPT taxa *Baetidae*, *Leptoceridae and Perlidae* in the Machangara River basin located in the southern Andes of Ecuador. With this objective, using generalized linear models (GLMs) habitat-suitability models were constructed, in which were analyzed the relation between the probability of occurrence of these pollution-sensitive macroinvertebrates families and physicochemical water quality conditions. The explanatory variables of the constructed GLMs differed substantially among the taxa, as did the preference range of the common predictors. In total, eight variables had a substantial influence on the outcomes of the three models. For choosing the best predictors of each studied taxa and for evaluation of the accuracy of its models, the Akaike information criterion (AIC) was used. The results indicated that the GLMs can be applied to predict either the presence or the absence of the invertebrate taxa and moreover, to clarify the relation to the environmental conditions of the stream. In this manner, these modelling tools can help to determine key variables for river restoration and protection management.

4.1 Introduction

The composition of the benthic macroinvertebrate communities can reflect the ecological quality of surface waters over time as pollution induces systematic shifts in community composition (Džeroski et al. 2000, Ambelu et al. 2010). Biological monitoring to assess river water health has been in use for more than a century. Nevertheless, in developing countries, this procedure was introduced and subsequently developed only recently (De Pauw et al. 2006). With biological samples, it is possible to predict the average values of chemical parameters over a period of time, when their accruing effects have been more pronounced in the biota (Džeroski et al. 2000).

For evaluation of the ecological water quality of rivers and lakes, some metrics based on taxonomic macroinvertebrate community composition have been developed, such as % Ephemeroptera—Plecoptera—Trichoptera (% EPT), % scrapers, taxa richness and Biological Monitoring Working Party (BMWP) (Rosenberg and Resh 1993, Ambelu et al. 2010). The latter index, which was developed in England and has been adapted to tropical countries, is a procedure to determine the water quality classes based on the score of sensitivity to organic pollution of the taxa found in water bodies (Junqueira and Campos 1998, Mustow 2002, Roldán Pérez 2003). The EPT taxa are often used because these families are the most sensitive orders and their richness is related to water quality, in particular, mainly to oxygen deficiency and environmental degradation (Hvitved-Jacobsen 1982, Thorne and Williams 1997, Lenat and Resh 2001).

Few studies have been conducted to understand the environmental preferences of macrobenthos in the high altitude Andes, in which some parameters have been identified as possible predictors of the presence of taxa. For instance, Ríos-Touma et al. (2011) found a relation between macroinvertebrate community composition and flow stability velocity and their variation in both dry and rainy seasons. Jacobsen and Marín (2008) principally analyzed the presence of the EPT taxa in relation to temperature and oxygen saturation. The same authors, in other research performed in Ecuador, found that the relation between the EPT taxa with the total of invertebrate fauna diminishes with the altitude (Jacobsen et al. 2003). Other research in the Andes indicated that the effect of organic pollution on macrobenthos was more evident during the dry season in comparison to the rainy season (Jacobsen 1998). However, other abiotic conditions that could have an effect on the environmental preference of macroinvertebrates, especially with regard to the sensitivity of EPT taxa were not investigated in these studies.

Prediction of the distribution of species based on biotic environmental conditions has been done through the modelling of running water. These models are now recognized as the core of predictive ecology and are powerful tools in conservation planning. The approach of these models is to predict the presence/absence or abundance of a species in relation to physicochemical variables and biotic and abiotic conditions collected in a specific habitat (Degraer et al. 2008, Zarkami 2008). Predicting the composition and distribution of macrobenthos communities in rivers is not a simple exercise due to the non-linear behavior of the ecosystem and the complexity of biotic and abiotic variables (Džeroski et al. 2000, Ambelu et al. 2010, Dominguez-Granda et al. 2011). For ecological modelling, techniques such as artificial neural networks (ANNs), Bayesian belief networks (BBNs), classification and regression trees (CTs and RTs), genetic algorithms (GAs), Logistic regressions (LRs) and support-vector machines (SVMs) have been used (Goethals 2005, Zarkami 2008). The developed models can be a comprehensive way to assess, with high reliability, the impact of an anthropogenic source in rivers, which can be helpful for decision support systems in river basin management (Ambelu et al. 2010).

The aim of this chapter is to determine which physicochemical parameters are related to the occurrence of three sensitive families, one of each order of the EPT taxa, in the Machangara River basin in Ecuador. To do so, environmental and biological variables were collected, which were the basis of the construction of three generalized linear models (GLMs). Furthermore, it is discussed the ecological relation between the selected predictors and their corresponding taxon, as well as the potential application and restriction of these models.

4.2 Data analysis

4.2.1 Model Species

Three macroinvertebrate families present in the river basin, which are pollution-sensitive based on the BMWP–Col, were selected. These taxa were *Baetidae* (Ephemeroptera), *Leptoceridae* (Trichoptera) and *Perlidae* (Plecoptera). *Baetidae* belong to the mayfly family and their tolerance score (TS) to pollution is 6. *Leptoceridae* are part of the caddisflies family and have a TS equal to 8. *Perlidae* are part of the stoneflies family and have a TS of 10.

4.2.2 Model Development, Selection, Validation and Optimization

In this chapter, the presence/absence of three sensitive macroinvertebrate taxa, described above, was associated with a physicochemical water quality condition, using a generalized linear model (GLM) with binomial distribution (cf. Section 2.3.1). Three different GLMs were constructed, thus one model for each family that can be used as input for the environmental suitability quantifications.

For starting the data exploration and analysis, boxplots linked to the presence and absence of a family in relation to the amplitude of physicochemical variables were constructed (Zuur et al. 2009). The water quality in the Machangara River was also analyzed, according to the BMWP-Col.

Collinearity between explanatory variables were computed with a non-parametric Spearman rank correlation coefficient, which is regularly used in ecology because this parameter makes no assumptions about linearity between the two variables (Zuur et al. 2009, Nguyen et al. 2014). Variables are not omitted for the construction of a model, when the Spearman correlation coefficient is lower than 0.80, since this means that no strong correlation is detected and no redundant variables are identified (Booth et al. 1994, Hering et al. 2006)

The effects of the input variables (i.e., abiotic variables) on GLM performance were assessed with a stepwise backwards selection procedure (cf. Section 2.3.1). This stepwise input variable selection procedure used in the construction process of the three models was tested to minimize the AIC. Each model had its own explanatory variables with which the presence or absence of an individual taxon was predicted. Homoscedasticity and normality were also checked based on the residual plots of the models. When the assumptions were violated, the models were discarded. All GLMs were fitted with an independent dataset and all statistical tests were performed at the 5% significance level. The GLMs and the boxplots were constructed in *R* (R-Core-Team 2015).

Finally, the model was validated with information collected in 14 points of the same basin in June of 2015 (dry season). Thus, the models developed with information collected during the rainy season were validated with a dataset collected during the dry season. In addition, the location of the samples sites was different between seasons. The accuracy of the validation set for each GLM was calculated as the pseudo R^2, which was determined as a relation between the correct predictions divided by the total points analyzed.

4.3 Results

4.3.1 Data Exploration

The flow had a wide variation with minimal values lower than 0.1 $m^3{\cdot}s^{-1}$ in small streams located in the upper area of the basin and maximum values of 23.1 $m^3{\cdot}s^{-1}$ in the lowest areas of the catchment, close to the city. The flow velocity, due to the mountainous terrain and the slope difference, varies from a very slow flow (0.07 $m{\cdot}s^{-1}$) in flat areas to high velocities (1.84 $m{\cdot}s^{-1}$) in steep lands (Fig. C1A). The water temperature was colder (9.1 °C) in the upper mountain regions than in the lower regions (13.4 °C) (Fig. C1A). The pollution from anthropogenic origin measured as BOD_5 (Fig. C1A), COD (Fig. C1B) and fecal coliforms (Fig. C1A) was low. The maximum concentration of these pathogens were 13 and 46 $mg{\cdot}L^{-1}$ and 5.4×10^5 MPN.100 mL^{-1} respectively. Values that were in the range of an analysis done by Esquivel et al. (2008), with more than 200 physicochemical samples taken on a quarterly basis by ETAPA-EP from 1984 to 2006. Because the fecal coliforms exhibited a wide variation, the analysis was executed on a logarithmic scale. For the most part, the pH remained in the basic zone, with three points with a level less than 7. The minimum value of this parameter was 6.33, while the maximum value was 8.36 (Fig. C1). The maximum value of the color was 40 HU, which was mainly vegetable in origin, due to the presence of moorland, wetlands and native vegetation (Fig. C1B). The conductivity in general was low, with values from 13.2 to 238 $\mu S \cdot cm^{-1}$ (Fig. C2C). The dissolved oxygen (DO) was high with a minimum of 6.7 $mg \cdot L^{-1}$ (Fig. C2D). This could be due to the high re-aeration capacity because of the high flow velocity and the low depth of the water in the river. When the Machangara River passes through the urban area, the maximum values of total solids (Fig. C2A) and turbidity (Fig. C2B) were registered: respectively at 190 $mg \cdot L^{-1}$ and 48.2 NTU. In general, the river had low depth, with only one place in the flat area located in the city, where this value reached more than 1.0 m (Fig. C1C). Other pollutants such as the organic nitrogen (Fig. C2D), the Nitrate+Nitrite-N, the Ammonium-N and the Phosphates (Fig. C1C) had low concentrations, with maximum values of 6.55 $mg \cdot L^{-1}$, 0.70 $mg \cdot L^{-1}$, 0.40 $mg \cdot L^{-1}$ and 0.55 $mg \cdot L^{-1}$ respectively. These points were registered where the land was used for pasture or crops. An overview of the measured physicochemical, microbiological and biological variables can be seen in Table 2.2.

Poor biological water quality was registered in the stretches after the dams of the two hydroelectric plants. The main cause of poor water quality here is hydropeaking. By contrast, good water quality results were found in the upstream locations of the basin where there is a low human presence. While, eight places located in the high land natural forest had moderate water quality, two locations in the reach flowing through the city registered poor biological water quality.

Regarding the fluctuations of the abiotic variables and their relation to the BMWP-Col, it was concluded that, when the flow velocity in the studied rivers was higher than 1.3 $m \cdot s^{-1}$, the poorest biological water quality was found (i.e., BMWP-class V: bad) (Fig. 4.1A). The highest concentration of the organic pollutants BOD_5 (Fig. C3A), COD (Fig. 4.1B), and fecal coliforms (Fig. C3B) was registered closest to the city, which agreed with the poorest biological quality (class V). Both COD and BOD_5, according to their minimum and maximum values (Fig. 4.1B and Fig. C3A), showed a wide range of variation when the poorest biological water quality was registered, while this range was lower to the other biological water classes. High flow velocities and increased levels of organic matter were detected in only one point close to the city; leading both parameters to register low biological water quality. In relation to the temperature, the biological water quality tended to be better with lower temperatures (Fig. 4.1C). It could be because in elevated regions, the anthropogenic pollution is lower and a protected area is present. The pH (Fig. C3C), the color (Fig. C3D), and the conductivity (Fig. 4.1D) do not have a visible impact on the BMWP-Class. However, when the BMWP-Col registered class three, the conductivity had the highest variation according its minimum and maximum values (Fig. 4.1D).

When the BMWP-Col is analyzed with the selected EPT taxa, it was determined that *Baetidae* was present in all biological classes (Fig. 4.2A). While, *Leptoceridae* was found when the BMWP-Col varied from class I to IV (Fig. 4.2B) and *Perlidae* was present in the highest biological classes (i.e., Class I, II and III) (Fig. 4.2C).

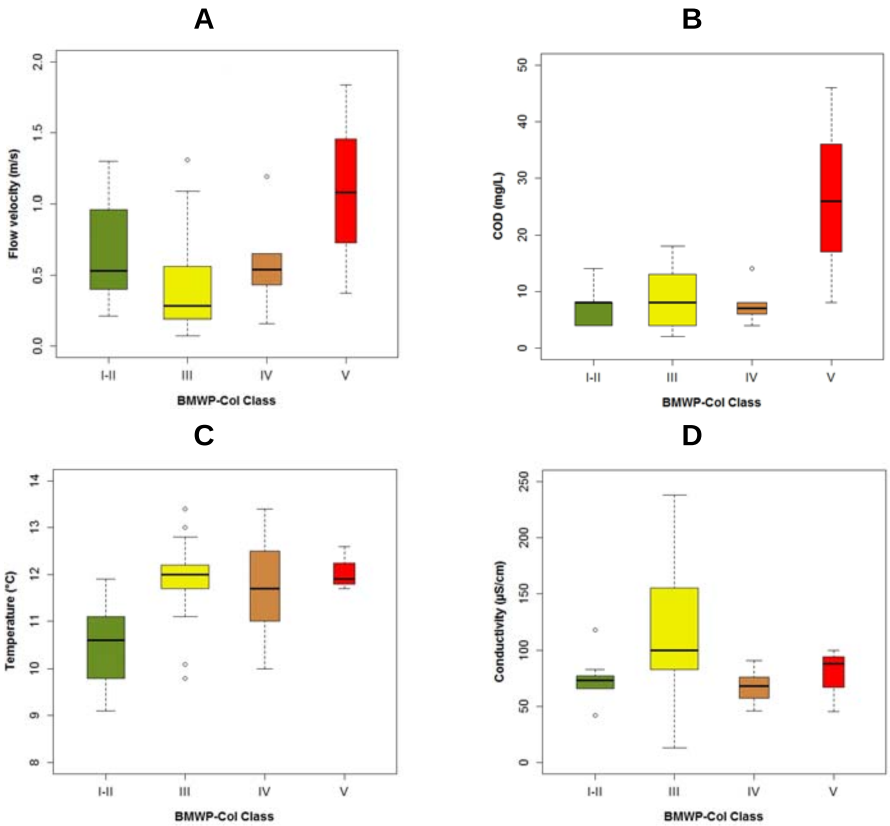


Fig. 4.1. Boxplots of the BMWP-Col class with the main explanatory variables of the three models: (**A**) flow velocity; (**B**) chemical oxygen demand (COD); (**C**) temperature and (**D**) conductivity.

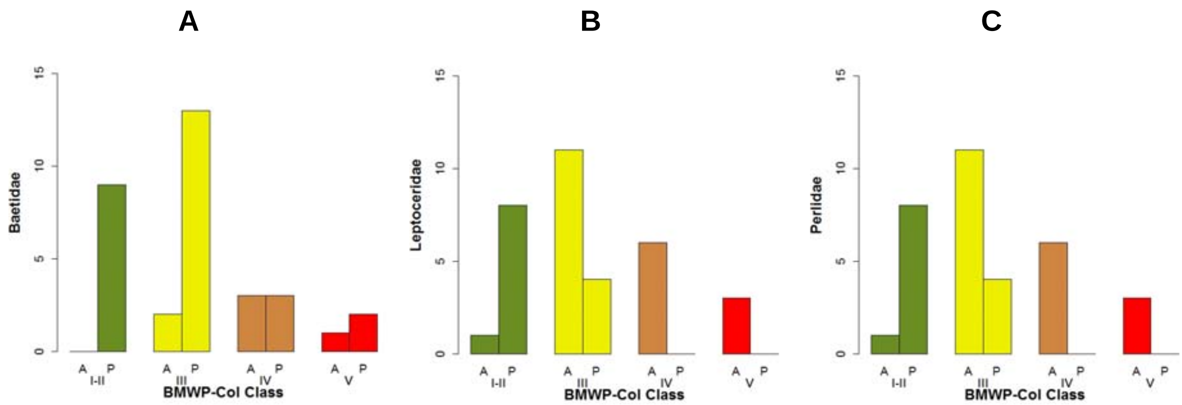


Fig. 4.2. Bar plots of the BMWP-Col class with the number of the samples in which the chosen macrobenthos were present (P) or absent (A): (**A**) *Baetidae*; (**B**) *Leptoceridae* and (**C**) *Perlidae*.

4.3.2 Correlation Analysis

Regarding the collinearity analysis (Table 4.1 and Table C1), no variables were omitted for the construction of the three models. Four cases, in which the absolute value of correlation coefficient is greater than 0.50 but less than 0.80, were detected.

For the *Leptoceridae* model, two values higher than 0.50 were found, COD with BOD_5 ($r = 0.61$, $p < 0.001$) and Log Fecal Coliforms with BOD_5 ($r = 0.57$, $p < 0.001$). In the *Baetidae* model, only one occurrence of the correlation coefficient higher than 0.50 was established, color with COD ($r = 0.63$, $p \leq 0.001$). One of the high values of correlation coefficient is color with BOD_5 ($r = 0.51$, $p = 0.002$), which is not in relation to any model.

Table 4.1. Spearman correlation coefficients of the explanatory variable used to construct the generalized linear models (GLMs).

Explanatory Variable	BOD_5	COD	Conductivity	Flow Velocity	Log Fecal Coliforms	pH	True Color (Color)	Water Temperature
BOD_5	1.0							
COD	0.6	1.0						
Conductivity	0.1	0.0	1.0					
Flow velocity	0.1	0.3	−0.3	1.0				
Log fecal coliforms	0.6	0.5	0.4	0.2	1.0			
pH	0.2	0.2	0.2	0.1	0.4	1.0		
True color (color)	0.5	0.6	−0.3	0.3	0.3	−0.1	1.0	
Water temperature	0.5	0.2	0.3	0.0	0.3	0.0	−0.0	1.0

4.3.3 Logistic Regression Models

Of the 16 physicochemical and microbiological variables measured in the Machangara River, only eight revealed a relation with the presence or absence in the three GLMs of *Baetidae*, *Leptoceridae and Perlidae.* These variables were flow velocity, water temperature, pH, color, conductivity, biological oxygen demand five (BOD_5), chemical oxygen demand (COD) and fecal coliforms (Table 4.2, Fig. C4–Fig. C6). However, the range of preferred physicochemical conditions differed between families in relation to their specific variables. The regression coefficients of Equation (2.1), *p*-values and goodness-of-fit measurements of the aforementioned models are presented in Table 4.2 and Table C2. For each of the families, the model selection procedure is summarized in Table C3–Table C5.

Table 4.2. Regression parameters: *p*-values and goodness-of-fit measurements of the models for predicting the presence of *Baetidae*, *Leptoceridae* and *Perlidae*.

Explanatory Variable	Regression Parameters	Baetidae		Leptoceridae		Perlidae	
		Coefficient	*p*-Values	Coefficient	*p*-Values	Coefficient	*p*-Values
	A	27.0	0.11	50.5	0.02	14.2	0.03
BOD_5	B1			−12.4	0.06		
COD	B2	0.8	0.10	−1.1	0.05		
Conductivity	B3	0.1	0.04	−0.1	0.06	0.1	0.04
Flow Velocity	B5			−12.3	0.04	2.3	0.17
Log Fecal Coliforms	B4			6.4	0.04	−1.9	0.02
pH	B6			−5.3	0.03		
Temperature	B7	−2.9	0.07			−1.4	0.03
True Color (Color)	B8	−0.4	0.10				
AIC:		22.5		26.5		34.5	
Pseudo R^2:		0.6		0.7		0.4	

For the family of *Baetidae* (TS: 6), the pseudo R^2 calculated for their GLM was around 0.6 and the AIC value was equal to 22.5, which was the lowest of the three constructed models for the three families. These taxa were present in 27 of the 33 sampled points. The effect of conductivity ($p = 0.04$), temperature ($p = 0.07$), color ($p < 0.10$) and COD

($p = 0.10$) were associated with the probability of the occurrence of *Baetidae* according to the constructed GLM (Table 4.2). Based on the constructed boxplots and the binomial model, the likelihood of the existence of *Baetidae* increases with lower temperatures (Fig. C4A and Fig. C7A). The effect of the color was similar, i.e., with a minor value, their possibility of presence increased (Fig. 4.3A and Fig. C4B). Concerning conductivity, *Baetidae* were likely to occur at higher values (Fig. C4D and Fig. C7C) compared with the other two analyzed taxa. Nevertheless, the maximum measured concentration of this variable was 238 $\mu S \cdot cm^{-1}$, which was relatively low. A higher probability of the existence of *Baetidae* was associated to the measured range (2 to 46 $mg \cdot L^{-1}$) of COD concentration, (Fig. C4C and Fig. C7B). The maximum registered value of this parameter corresponded to water with relatively low pollution. The plots of the residuals vs. fitted and scale-location of the model can be seen in Fig. C10A and Fig. C10B, respectively.

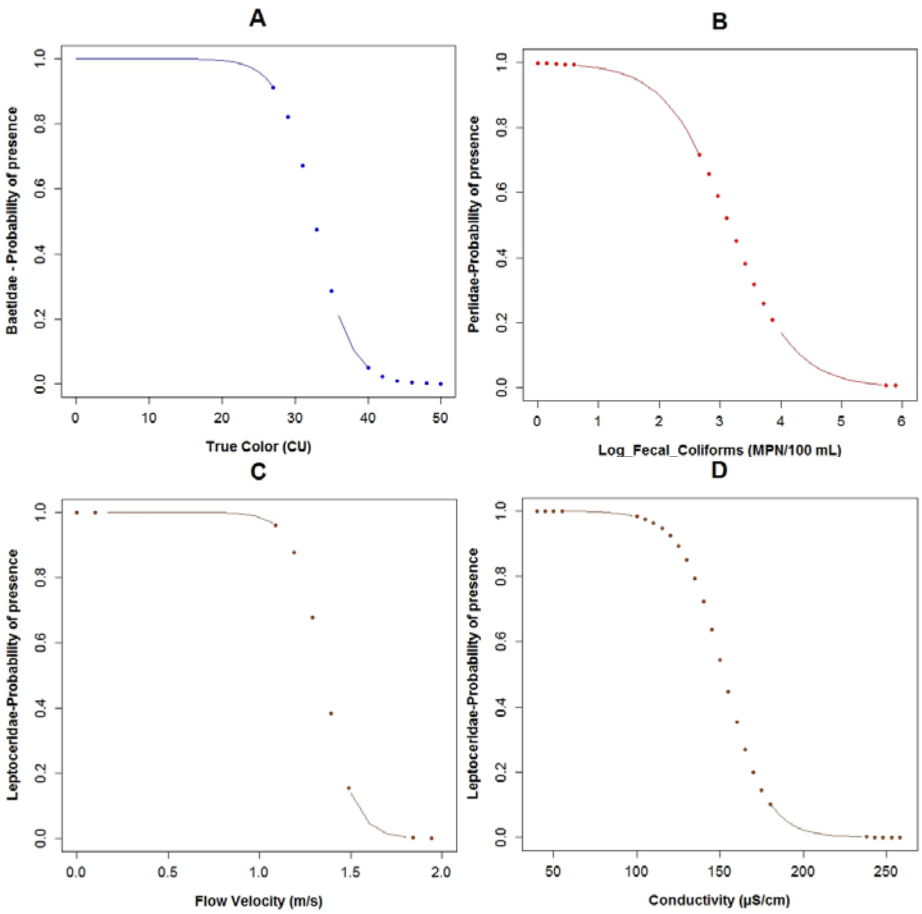


Fig. 4.3. The probability of chosen taxa being present in relation to an explanatory variable: (**A**) *Baetidae* with color; (**B**) *Perlidae* with log fecal coliforms; (**C**) *Leptoceridae* with flow velocity; (**D**) *Leptoceridae* with conductivity. (The points in the curves indicate extrapolation outside the observed physicochemical variables range).

4

Regarding *Leptoceridae* (TS: 8), their GLM had the highest pseudo R^2 of the three analyzed models for the three taxa, with a value of 0.7 and the AIC value was equal to 26.5. This family was present in six of the 33 sampled points. The pH (p = 0.03), flow velocity (p = 0.04), fecal coliforms (p = 0.04) expressed in log scale, COD (p = 0.05), BOD_5 (p = 0.06) and conductivity (p = 0.06) are the variables of the GLM related to the probability of the occurrence of this family (Table 4.2). The boxplots of *Leptoceridae* and their GLM reveals that the probability of the presence of these taxa is higher with low values of fecal coliforms (Fig. C5A and Fig. C8A), low concentrations of COD (Fig. C5E and Fig. C8B), and BOD_5 (Fig. C5C and Fig. C8C). The probability of the existence of the aforementioned taxa is higher with neutral pH (Fig. C5D and Fig. C8D) and when the flow velocity is lower than one meter per second (Fig. 4.3C and Fig. C5B). These taxa tend to be present when the conductivity is low (Fig. 4.3D and Fig. C5F). The plots of the residuals vs. fitted and scale-location of the model can be seen in Fig. C10C and Fig. C10D, respectively.

The *Perlidae* (TS: 10) GLM model was characterized by the lowest pseudo R^2 value (0.4) and the highest AIC value (34.5) of the three constructed models for the three families. These taxa were present in 12 of the 33 sampled points. The probability of the occurrence of *Perlidae* in relation to the constructed GLM is associated to fecal coliforms (p = 0.02) expressed in log scale, temperature (p = 0.03), conductivity (p = 0.04) and flow velocity (p = 0.17) (Table 4.2). Regarding the binomial model of *Perlidae* and their boxplots, the possibility of the presence of this family is higher with less concentration of fecal coliforms (Fig. 4.3B and Fig. C6B). When the temperature (Fig. C6A and Fig. C9A) was lower, the opportunity of occurrence of the aforementioned taxa increased. *Perlidae* also prefer flow velocities with values less than 1.5 $m \cdot s^{-1}$ (Fig. C6D and Fig. C9B) and conductivity below 180 $\mu S \cdot cm^{-1}$ (Fig. C6C and Fig. C9C). The plots of the residuals vs. fitted and scale-location of the model can be seen in Fig. C10E and Fig. C10F, respectively.

The studied families: *Baetidae*, *Leptoceridae and Perlidae,* had one common variable, the conductivity, which always was below 250 $\mu S \cdot cm^{-1}$. COD was a conjoint variable between *Baetidae* and *Leptoceridae*, albeit the amplitude of this predictor was lower when the sensitivity of the taxa was higher. *Leptoceridae* and *Perlidae* had two other mutual variables: fecal coliforms and flow velocity. For the first variable, the range was minor when the sensitivity of the taxa was greater, while for the second predictor, the value in both cases was less than 1.4 $m \cdot s^{-1}$. A common variable between the GLMs of *Baetidae* and *Perlidae* was the temperature, which was inferior when the sensitivity of the taxa was superior.

When the GLMs were validated with an independent data set, the accuracy of the models measured as pseudo R^2 was 0.9 for *Baetidae*, 0.9 for *Leptoceridae* and 0.4 for *Perlidae*.

4.4 Discussion

4.4.1 Analysis of the Chosen EPT Taxa in Relation with the BMWP-Col

Although the average values of measured pollutants were low in eight points located by the forest in the upper area of the basin, the biological class given for BMWP-Col was moderate. Similar findings were found in high land streams in the Andes in South America and in the Rwenzori Mountains in Africa (Beadle 1974, Jacobsen and Marín 2008, Kasangaki et al. 2008). This could be due to environmental stress caused by natural factors such as the disturbance of stream sites due to poor habitat conditions, the impact of heavy rains in the rainy season (Jacobsen and Marín 2008), and heavy shading, or due to hydropeaking as a result of dam operation. Other causes could be the water physicochemical characteristics in the forest sites, such as low conductivity, low turbidity, high dissolved oxygen concentration, which in combination with heavy shading could influence the low primary production (Acreman and Ferguson 2010).

4

The presence of *Baetidae* in all the BMWP-Col classes could be due to the fact that the maximum measured concentration of BOD_5 was 13 $mg \cdot L^{-1}$, a value that is relatively low, and the minimum DO concentration was high with a value of 6.65 $mg \cdot L^{-1}$. This family was found in Thailand with concentrations of BOD_5, DO and conductivity of 7 $mg \cdot L^{-1}$, 1.1 $mg \cdot L^{-1}$ and 377 $\mu S \cdot cm^{-1}$ respectively. In Ghana, these taxa were present with a BOD_5 of 2 $mg \cdot L^{-1}$, DO of 4.1 $mg \cdot L^{-1}$ and a conductivity of 1250 $\mu S \cdot cm^{-1}$ (Thorne and Williams 1997), while in Turkey, *Baetidae* were found when BOD_5 was 8.8 $mg \cdot L^{-1}$ (Ogleni and Topal 2011).

When *Leptoceridae* were present, the BMWP-Col class varied from I to IV, despite low pollution registered, with a maximum BOD_5 of 0.60 $mg \cdot L^{-1}$ and a minimum DO of 6.7 $mg \cdot L^{-1}$. *Perlidae* were present when the BMWP-Col was registered from Class I to III, the minimum DO was 7.0 $mg \cdot L^{-1}$ and the maximum BOD_5 was 0.50 $mg \cdot L^{-1}$. Other factors such as disturbances of stream caused by the rainy season, or by the daily operation of the two dams, could have induced the lower BMWP-Col class.

4.4.2 Analysis of the Explanatory Variables in Relation to Response Variables

Our results demonstrate that GLMs can be used to select physicochemical variables that best predict the presence of the EPT taxa. These orders were chosen because those models describing environmental preference conditions of sensitive families are more reliable and valid than models for non-sensitive taxa (Forio et al. 2016). For example, Forio et al. (2016) showed that the model constructed to predict the occurrence of the *Leptophlebiidae* (TS: 10) taxa, which belong to the Ephemeroptera order, was reliable, while, for the prediction of the *Chironomidae* (TS: 2), which belong to the Diptera order, the model was not reliable.

Conductivity was a common variable between the three analyzed taxa (*Baetidae*, *Leptoceridae and Perlidae*). This variable represents the natural mineral content of the water from both inorganic and organic origin (D'Heygere et al. 2003). In the first case, inert material is dragged because of precipitation and local geology, as well as inorganic pollutants from different anthropogenic sources that are leached or discharged into water bodies. However, in the second case, the organic origin is mainly due to wastewater discharges (Gabriels et al. 2007, Schwarzenbach et al. 2010). In agreement with the models developed in this chapter, conductivity has been established as one of the most critical variables to predict macroinvertebrates presence in rivers, in both tropical and temperate countries (D'Heygere et al. 2003, Lock and Goethals 2013, Everaert et al. 2014). In this way, the caddisflies and stoneflies orders, from which *Leptoceridae* and *Perlidae* originate respectively, were reported to be present when the conductivity was relatively low (Lock and Goethals 2008, 2012). Furthermore, D'Heygere et al. (2003) indicated that the wide range of variation of this variable expresses the high diversity existing between rivers and their stretches.

Similarly, Dissolved Oxygen (DO) is a variable that in most cases is present to predict the occurrence of macroinvertebrates (D'Heygere et al. 2003, Dedecker et al. 2004, Holguin-Gonzalez et al. 2013a). Nonetheless, DO was not a variable of the three GLMs in this research, an observation that was not expected in the research hypothesis of this chapter. The main cause could be that the lowest measured concentration of DO in the field was 6.65 $mg \cdot L^{-1}$ (>85% oxygen saturation), a value that was enough for the presence of *Baetidae*. For the occurrence of *Leptoceridae* and *Perlidae*, higher concentrations of DO were necessary, values that were greater when the family was more sensitive. That is, for *Leptoceridae*, the minimum DO concentration was 6.67 $mg \cdot L^{-1}$, while for *Perlidae* the lowest concentration of this variable was 7.0 $mg \cdot L^{-1}$. Additionally, mountain rivers are likely to maintain relatively high oxygen saturation due to their high velocity and turbulence. Consequently, these rivers have better aeration and less sedimentation of oxygen-consuming materials thus being more resistant to organic pollution (Jacobsen 1998). Regarding Dissolved Oxygen, Connolly et al. (2004) mentioned that mayflies populations declined dramatically when the DO levels were less than 20% of saturation. Moreover, a finding in the highlands of Ecuador consistent with the analysis presented in this chapter, established that no EPT taxa were present in tributaries when the oxygen saturation was lower than 80% (Jacobsen et al. 2003).

Flow velocity is another important variable that has been analyzed by some authors. In this research, the aforementioned variable is part of the two predictable models for *Leptoceridae and Perlidae* whereas it is not a factor for the *Baetidae* model. Interestingly, Holguin-Gonzalez et al. (2013a) reported the flow velocity as a predictor of the presence of macroinvertebrates in Colombia. In the same way, Ríos-Touma et al. (2011) in their research, in the high altitude tropical streams in Ecuador, found that flow velocity had influence in abundance, in communities such as *Leptoceridae.* However, Ríos-Touma et al. (2011) reported that this variable is less important for macrobenthos with faster development times such as *Baetidae,* which is less susceptible to hydraulic disturbance. A similar finding in Zimbabwe for the Trichoptera order, showed a strong relation between their occurrence and abiotic parameters such as flow velocity and average depth (Chakona et al. 2009).

The temperature variation between high altitudes despite short distances to reach them is considerable. In addition, the hourly and daily temperatures at one altitude fluctuate substantially in the tropical Andes Mountains (Ríos-Touma et al. 2011). Furthermore, due to lower water temperature, the solubility of oxygen increases leading to elevated DO concentrations in the water. At higher elevations, however, the atmospheric partial pressure of oxygen diminishes with ascension in altitude, countering the effects of lower temperature (Jacobsen 1998). The aforementioned temperature is part of two GLMs (*Baetidae* and *Perlidae*) and it is negatively correlated with the possible presence of these two taxa. A similar correlation was found for the prediction of the EPT taxa in the tropical Andean region of Bolivia (Jacobsen and Marín 2008) or for the macrobenthos assemblage in California (Hawkins et al. 1997).

The significant variables of the three GLMs that were related to organic pollution were COD, BOD_5 and fecal coliforms. The first variable was in relation to the possible presence of *Baetidae* and *Leptoceridae,* while BOD_5 was only associated with the probable occurrence of *Leptoceridae* and fecal coliforms related to the possible existence of *Baetidae* and *Perlidae*. Outcomes were consistent with Jacobsen (1998), who specified that the effect of organic pollution on the macrobenthos in Ecuadorian high land tributaries was the same as rivers at higher latitudes. In this way, Jacobsen (1998) and Ríos-Touma et al. (2011) reported in their research in the Andes of Ecuador that Plecoptera, in which *Perlidae* originates, was present in pristine conditions and unpolluted places. With regard to *Leptoceridae*, the previously mentioned authors described that this family was present in only slightly polluted places. While, with regard to *Baetidae*, Jacobsen (1998) expressed that these taxa were not found in severely polluted streams. Macrobenthos are known to be affected by organic pollution for two reasons: the first is dissolved oxygen reduction and the second is due to alteration of the substratum and loss of available food sources (Hynes 1960). Furthermore, a study performed in the Itambi River, located in the northern Andes of Ecuador, evaluated the impact of organic pollution. The study concluded that the number of macroinvertebrate species was reduced when the concentration of BOD_5 increased with a subsequent decrease in DO and vice versa (Burneo and Gunkel 2003). In the samples subject of this research, it was found that when *Perlidae* were present the maximum COD and BOD_5 were equal to 14 $mg \cdot L^{-1}$ and 0.5 $mg \cdot L^{-1}$ respectively. The relatively low concentration of COD and BOD_5, could be the main reason these variables were not present in the *Perlidae* Model. In either case, when one of the three families was present, the maximum concentration of COD was equal to 46 mg·L^{-1}, an amount registered for *Baetidae*. For more sensitive taxa, lower values of COD were registered for their presence. These findings are in line for the EPT taxa, which are considered indicators of relatively clean water due to their sensitivity to pollution (Rosenberg and Resh 1993, Džeroski et al. 2000, Jacobsen and Marín 2008).

Two other variables, color and pH, had respective relationships to *Baetidae* or *Leptoceridae*. On the one hand, color had a negative correlation with the model to predict the probable presence of *Baetidae*. The range of this predictor, when these taxa were found, varied from 0 to 27 Hazen units (HU). The color in natural water consisted mainly of the generic humic and fulvic fraction of dissolved organics (Bennett and Drikas 1993), and its intensity could be increased in direct relation to the amount of precipitation and runoff (Haaland et al. 2010). The humic substances are organic acids and their accumulation and dissolution in water are associated with a reduction of pH (Beadle 1974, Dallas and Day 2007). The pH also had a negative correlation as a possible predictor of the presence of *Leptoceridae*, which was found when pH was between 6.33 and 8. Similar values (4.6 to 7.9) were found when *Leptoceridae* were present in research done in the high-altitude streams in Uganda (Kasangaki et al. 2008). In addition, the composition and abundance of macroinvertabrate taxa have been shown to have a relationship with pH (Feldman and Connor 1992) and subtle differences in this parameter may explain the differences in macrobenthos assemblages (Dallas and Day 2007). The potential of hydrogen has been used to predict the presence of some families of macrobenthos, such as *Baetidae*, *Hydroptilidae* and *Simuliidae* (Kasangaki et al. 2008, Ambelu et al. 2014, Everaert et al. 2014).

The application of the three GLMs to find the probable presence of the three families studied (i.e., *Baetidae*, *Leptoceridae and Perlidae*) is defined by the measured range of the predictors found in this research. Regarding this topic, some authors have written that the preferential conditions of a family varies from place to place, in connection with abiotic settings (Jacobsen and Marín 2008, Everaert et al. 2014). Hence, the three constructed models could not be simulated in different amplitude of the predictors that were evaluated in this investigation.

4.4.3 Model Performance

Three GLMs for three different EPT taxa (i.e., *Baetidae*, *Leptoceridae and Perlidae*) were built in this chapter using a generalized linear model. GLMs have the advantage of directly establishing and reporting the relative importance of each variable in searching for the biotic integrity (Damanik-Ambarita et al. 2016). Furthermore, a stepwise discriminant procedure to select the most significant variables based on the AIC selection criteria applied in this research, has been shown to be effective in the prediction of species distribution (Thuiller 2003). However, the variables chosen for the developed models were not necessarily the only ones that were important.

Variables that were not selected could be due to a correlation with another variable or with a set of variables (Ambelu et al. 2010), or that they were less important for the model than variables that were selected.

Because the data sample size (33 points) was relatively small (<100 points) (Stockwell and Peterson 2002, Vaughan and Ormerod 2005), the binary data that represents their presence and absence was applied for the construction of the GLMs. With regard to small sample sizes, Stockwell and Peterson (2002) concluded that the accuracy of samples of ten points was 90% and when the size increased to 50 points their accuracy was near maximal. In addition, Logistic regression models that were applied to different small sample sizes showed a stable goodness-of-fit in their predictable capability (Holguin-Gonzalez et al. 2013b, Yu and Abdel-Aty 2013). Despite the fact that the basin is regulated and taking into account the samples were only taken in the rainy season, binary construction was applied for the model development. This kind of development based on one season gives the best results, however, it does not allow for insights in to seasonal variations related to the abundance of macrobenthos (Everaert et al. 2014). Moreover, the probability of the existence of taxa based on their presence or absence was estimated with a Logistic regression, which is more suitable and often used for these types of predictions (Pearce and Ferrier 2000, Manel et al. 2001, Zuur et al. 2009).

The suitability of the developed models in the first instance is in the range in which the eight explanatory variables were measured (Fig. C1 and Fig. C2). For *Baetidae*, the presence or absence of this taxon as a result of their GLM, could be inferred in lower or higher values of color and temperature. However, it is not recommend extrapolating for the possible presence of this family when the concentrations are greater than the measured range for the COD (46 $mg \cdot L^{-1}$) and conductivity (238 $\mu S \cdot cm^{-1}$). The presence or absence of *Leptoceridae* according to the developed GLM, could be extrapolated by flow velocity, conductivity, COD, BOD_5 and pH. Nonetheless, it is not advice to infer the possible presence of this family when fecal coliforms are higher than the measured range (1.7×10^5 MPN.100 mL^{-1}). Regarding the *Perlidae* model, the temperature and fecal coliforms could be extrapolated in smaller and greater measured values for determination of their presence or absence. However, this model must not be extrapolated without prior determination of the presence of this taxa in higher values of flow velocity (1.30 $m \cdot s^{-1}$) and conductivity (177.9 $\mu S \cdot cm^{-1}$).

According to Mac Nally (2000), the typical R^2 found in 100 cases analyzed was around 50% with $p > 0.5$ and R^2 value diminished to 25%, when only the significant variables were retained in the models. The author mentioned also specified that much simpler models with less variables produced typical R^2 around 30% with $p < 0.001$.The goodness-of-fit of the three GLMs, which was also measured with the correlation coefficient pseudo R^2, has shown that the *Leptoceridae* model had the highest value (67.6%), followed by the *Baetidae* (60%) and the *Perlidae* (43.5%) GLMs, a common adjustment range found in ecological models. When the accuracy of the model was assessed with an independent validation dataset, the pseudo R^2 was similar for both the *Baetidae* and the *Leptoceridae* (86%), while for *Perlidae* it was the lowest with an analogous value that was obtained with the training data set. In the latter value, it is not clear why this was the lowest.

4.5 Conclusions

Three generalized linear models (GLMs) were built in order to understand the physicochemical water quality variables that determine the relation to the presence of three selected EPT taxa. Eight variables, identified during the stepwise selection procedure, showed a clear relation to the probable occurrence of the analyzed families. Each taxon had its own explanatory variables, of which conductivity was a unique common term between the GLMs of the three families. The fit of the GLMs was measured with the Akaike information criterion (AIC), the pseudo R^2, as well as, an independent validation data set. The pseudo R^2 varied from 43.5% to 67.6%, values that are common for ecological models. Therefore, the GLMs could be used as tools to predict changes in the biological quality of the Machangara River.

Chapter 5. Model-Based Analysis of the Potential of Macroinvertebrates as Indicators for Microbial Pathogens in Rivers

Adapted from:

Rubén Jerves-Cobo, Gonzalo Córdova-Vela, Xavier Iñiguez-Vela, Catalina Díaz-Granda, Wout Van Echelpoel, Felipe Cisneros, Ingmar Nopens and Peter L. M. Goethals. (2018). Model-Based Analysis of the Potential of Macroinvertebrates as Indicators for Microbial Pathogens in Rivers. Water 2018, vol. 10, p. 375; doi:10.3390/w10040375. ISSN 2073-4441.

Abstract

The quality of water prior to its use for drinking, farming or recreational purposes must comply with several physicochemical and microbiological standards to safeguard society and the environment. In order to satisfy these standards, expensive analyses and highly trained personnel in laboratories are required. Whereas macroinvertebrates have been used as ecological indicators to review the health of aquatic ecosystems. In this chapter, the relationship between microbial pathogens and macrobenthic invertebrate taxa was examined in the Machangara River located in the southern Andes of Ecuador, in which 33 sites, according to their land use, were chosen to collect physicochemical, microbiological and biological parameters. Habitat-suitability models based on decision tree models (DTMs) were used to generate rules that link the presence and abundance of some benthic families to microbial pathogen standards; a pragmatic way to determine poor and risky water quality conditions. The aforementioned DTMs provide an indirect, approximate, and quick way of checking the fulfillment of Ecuadorian regulations for water use related to microbial pathogens. Thus, *Baetidae* and *Scirtidae* could be identified, as negative indicators meaning that their abundance with at least four and five individuals in the sampling hand-net method could be a signal of poor water quality with hazardous use for drinking. The models built and optimized with the WEKA package, were evaluated based on both statistical and ecological criteria to make them as clear and simple as possible. As a result, two different and reliable models were obtained, which could be used to predict proxy indicators in preliminary assessment of pollution of microbial pathogens in rivers. The DTMs can be easily applied by staff with minimal training in the identification of the taxa selected by the models. The presence and abundance of selected macroinvertebrate taxa in conjunction with the decision trees can be used as a screening tool to evaluate sites that require additional follow up analyses to confirm whether microbial water quality standards are met.

5.1 Introduction

The most frequent health risk related to the ingestion of water is associated with microbial contamination by human or animal feces, which is a source of pathogenic bacteria, viruses, protozoa and helminthes (Gofti-Laroche et al. 2003, WHO 2004). Pathogens are introduced in rivers via point and non-point sources, and their autochthonous growth is stimulated by nutrients brought from the aforementioned sources (Mallin et al. 2006). The health risk increases when untreated wastewater from urban sewage systems (point source) is directly discharged into water bodies, potentially causing large outbreaks of waterborne diseases (Arnone and Walling 2007). In addition, water from rivers and lakes has off stream uses such as drinking water or irrigation, and instream uses such as recreational activities with primary contact (e.g., swimming). Therefore, water quality control must always be of paramount importance (Oliver 1984).

The indicators often used to verify microbial contamination of water in developed countries are: total coliforms, and fecal coliforms and/or *Escherichia coli* (Fewtrell et al. 2001, WHO 2003). Likewise, in many tropical countries, the assessment of running water quality is predominantly made by using physicochemical methods. However, most of the methods for determining physicochemical and microbiological parameters require expensive laboratory analyses that in the majority of developing countries, do not allow for the establishment of national rigorous monitoring programs of water bodies due to limited technical and financial resources. For those reasons, the development of cost-effective water monitoring programs is essential (Dominguez-Granda et al. 2011), and must include techniques for measuring microbial water quality.

The biological methods for monitoring river water health have evolved over more than a century. For example, benthic macroinvertebrates are used to assess the water quality over time, because they respond to both physicochemical changes and hydro-morphological variations within streams and rivers (De Pauw et al. 2006, Gabriels et al. 2010). Physicochemical and microbiological parameters provide limited water quality information at a specific point in time (Džeroski et al. 2000, De Pauw et al. 2006). In contrast, biological samples can also predict average values of chemical parameters when their cumulative effects have been more pronounced in the biota over a period of time preceding the biological sampling (Džeroski et al. 2000). As such, the use of bioindicators in water quality assessment for streams has been integrated into the European Water Framework Directive (Griffiths 2002). In developing countries, biological river assessment was introduced and subsequently developed only recently (De Pauw et al. 2006), based mainly on adaptation of the English Biological Monitoring Working Party (BMWP) (Junqueira and Campos 1998, Mustow 2002, Roldán Pérez 2003).

Fecal coliform (FC) concentration has been modeled using both deterministic and stochastic methods. The deterministic models focused on understanding the die-off variation of fecal coliforms in relation to temperature, and changes under kinetics conditions (i.e., transportation) such as the velocity along the rivers (Wilkinson et al. 1995). Alternatively, stochastic models have been used to obtain the relationship between fecal coliform and physicochemical (Mahloch 1974) or microbiological (Ansa et al. 2012) variables, or timing variation during a rainfall (Kay and McDonald 1983). Negative correlation between FC concentrations and macroinvertebrate diversity (Shannon-Wiener diversity index) was observed in ponds (Ansa et al. 2012) and increase of an specific negative indicators (*Chironomidae*) could be observed in relation to effluent wastewater outfalls (Hynes 1960).

On the other hand, the assessment of habitats and the determination of the relation between the presence of an organism and environmental variables has been done through the modelling of running waters based on ecological, physicochemical and microbiological parameters. These modelling techniques have allowed for the handling of the non-linear behavior of the ecosystem, obtaining models with a high reliability (Hoang et al. 2010b, Ambelu et al. 2014, Jerves-Cobo et al. 2017). In this way, the FC has been associated as one of the explanatory variables describing the presence or absence of some taxa of macroinvertebrates (Holguin-Gonzalez et al. 2013a, Acosta and Hampel 2015, Jerves-Cobo et al. 2017). Machine learning with different modelling techniques, such as classification trees (CTs) combine reliable classification predictions with transparency, and have been proven to be effective to assess running waters (Goethals et al. 2006, Ambelu et al. 2010). The CTs provide good modelling techniques as they focus on the presence/absence or abundance of macroinvertebrate taxa (family or species) in relation to a specific impact or a disturbance in the streams (Džeroski et al. 2000, Goethals 2005, Dakou et al. 2007, Ambelu et al. 2010, Hoang et al. 2010b). Consequently, considering the described correlations between fecal coliform presence and macroinvertebrate diversity (Ansa et al. 2012, Holguin-Gonzalez et al. 2013a, Acosta and Hampel 2015, Jerves-Cobo et al. 2017), compliance to regulatory standards can be simulated based on the prevailing macroinvertebrate community structure by training classification trees on combined observations of fecal coliforms and macroinvertebrates, thereby acting as a proxy indicator for fecal coliform contamination.

In this chapter, with the environmental and biological variables collected in the Machangara River in Ecuador between February and March of 2012, two decision tree models (DTMs) were developed as indicator tools to check the compliance to three of the Ecuadorian microbial water quality standards associated with fecal coliforms. The construction of the DTMs was based on the presence and abundance of macroinvertebrates in the Machangara River basin. The models were built based on statistical adjustments and ecological criteria. For model optimization, statistical techniques were used, such as the elimination of false positives (*FP*) achieved by applying weights as well as the minimum confusion entropy from the models. The final two DTMs were validated with datasets collected in July of 2015 and March of 2016.

5.2 Data analysis

5.2.1 Ecuadorian Water Regulation in Relation to Water Use

The Ecuadorian government has regulations regarding the water quality in relation to water use (MAE-Ecuador 2015). The standard norm set a value limit for different parameters in relation to particular water usage, giving three thresholds to regulate the concentration of fecal coliforms with regard to water use (Table 5.1). The most stringent microbial water quality standard for fecal coliforms is applied for recreational water use with primary contact (Table 5.1). The least stringent microbial water quality standard for fecal coliforms is for raw (untreated) water used for drinking water before receiving non-conventional treatment (Table 5.1). Non-conventional treatment methods include slow sand filtration and multi-stage filtration, which is recommended for small towns that need flows less than 8 L/s and a population <5000 people or whose town needs flows up to 21 L/s and a population <12,000 people (Sánchez et al. 2007). The intermediate microbial water quality standard for fecal coliforms is for agriculture (Table 5.1).

5

Table 5.1. Ecuadorian Water Quality Regulation for fecal coliforms (MAE-Ecuador 2015).

Regulations	Water Used for	Fecal Coliforms Limited Value MPN.100 mL^{-1}
Recreational	Recreational with primary contact	≤200
Agriculture	Agriculture and livestock	≤1000
Raw water	Raw water previous to non-conventional treatment [a]	≤2000

[a] Conventional treatment refers to chemical addition, rapid mixing, flocculation and sedimentation; MPN = Most probable number.

5.2.2 Model Development

Decision tree models (DTMs) were developed to predict the fecal coliforms regulation fulfillment according to the water uses (Table 5.1), and were expressed as three discrete levels. In this research, the attributes or explanatory variables of the DTMs were the presence/absence or abundance of macroinvertebrates taxa that were observed in at least three sampled points (Table A4 and Table A5). The discrete response variables were in fulfillment of the three microbial water quality standards for fecal coliforms, which were measured as most probable number per 100 mL (MPN.100 mL^{-1}). The classification trees are robust techniques that can deal with small datasets (Forio et al. 2016) less than 50 data points (Stockwell and Peterson 2002), particular to the case of this study, in which the dataset is composed of the results of 33 different sites. In addition, with a small dataset the accuracy of the classification trees models is higher than other techniques such as Logistic regression models (Yu and Abdel-Aty 2013). All observations were included to construct the models, because classification trees are not sensitive to outliers (Moisen 2008).

In this chapter, the machine learning software, Waikato Environment for Knowledge Analysis (Weka) (Witten and Frank 2005), and its package J4.8 decision tree classifier that is a Java re-implementation of C4.5 (Quinlan 2014) were used for inducing classification trees and creating a prediction model. The model training and validation were performed with three, five, ten-fold (k-fold) cross-validation (three, five, 10, k fcv) in which the records are randomly split into k equally-sized subsets. In each set, k-1 subgroups are used as the training set and the k-th that remains is run as the test set. This process is repeated k times and each subset is used as the test set exactly once (Goethals 2005). The expansion of the tree is stopped with the pruning process, which gives to every leaf, a minimum number of instances to allow branching. With the aim to improve this process, two pruning confidence factors (PCF) were employed: 0.25, which is the default value, and 0.1. With a small dataset, lower cross-validation values can result in more robust models, but with a relatively low performance (Goethals et al. 2007). Table 5.2 and Table D1 in Appendix D show the settings for the eight models obtained with three, five, 10 fcv and 66% of data as a trained set, as well as a PCF of 0.1 and 0.25, before optimization.

Table 5.2. Model reliability evaluation from the model development phase that were based on the J4.8 algorithm (before optimization): correctly classified instances (CCI), Kappa statistics and overall confusion entropy of a confusion matrix (CEN).

Model No.	FCR [a]	Model Settings		Model Outcomes			
				CCI [c] (%)	Kappa Statistics	Number of Leaves	CEN [d]
		J4.8	PCF [b]	CCI ± sd	Kappa ± sd		CEN ± sd
1[e]ap[f]1[g]	Recreational [h]	3, 5 and 10 fcv [i]	0.25	40.4 ± 3.5	−0.2 ± 0.1	6	1.0 ± 0.0
1ap2	Recreational	3, 5 and 10 fcv	0.10	48.5 ± 3.0	−0.1 ± 0.1	2	1.0 ± 0.0
1a1	Recreational	3, 5, 10 fcv and 66%tr	0.25	70.5 ± 1.5	0.4 ± 0.1	5	0.8 ± 0.0
1a2	Recreational	3, 5, 10 fcv and 66%tr	0.10	70.5 ± 1.5	0.4 ± 0.1	4	0.8 ± 0.0
2ap1	Agriculture	3, 5 and 10 fcv	0.25	66.7 ± 0.0	0.2 ± 0.0	4	0.9 ± 0.0
2ap2	Agriculture [j]	3, 5 and 10 fcv	0.10	69.7 ± 3.0	0.2 ± 0.0	3	0.8 ± 0.0
2a1	Agriculture	3, 5, 10 fcv and 66%tr	0.25	86.4 ± 8.0	0.7 ± 0.2	3	0.5 ± 0.2
2a2	Agriculture	3, 5, 10 fcv and 66%tr	0.10	77.3 ± 16.9	0.4 ± 0.4	3	0.7 ± 0.3

Mean and standard deviations of CCI, Kappa statistics and CEN were derived from k-fold cross-validation. In the case of two or more models that had the same DTM, these parameters were obtained from a k-fold cross-validation of all related models. [a] FCR = Fecal coliform regulation; [b] PCF = Pruning confidence factor; [c] CCI= Correctly classified instances; [d] CEN = Overall confusion entropy of a confusion matrix; [e] fecal coliform regulation: 1 for recreational and 2 for agriculture; [f] The kind of database: ap = absence/presence, a = abundance; [g] The number of model with different value of PCF; [h] The short name of FCR; [i] fcv = folds cross-validation; [j] Models obtained from agriculture regulation could be applied to check the raw water fecal regulation.

5.2.3 Model Optimization

Optimization was achieved by adding costs with a cost-sensitive classifier (CSC) tool with the J4.8 algorithm in the WEKA software. The CSC process gives new weights in training instances according to the total cost assigned to two kinds of errors: false positives (*FP*), which are known as type II errors, and false negatives (*FN*), which are identified as type I errors, with the least expected misclassification cost rather than the most likely one (Witten and Frank 2005). The differences between the cost sensitive classifier (CSC) with the J4.8 algorithm is the initial setting in this process in which different weights to false positive have been given in the cost matrix that seeks to minimize the number of type II (i.e., *FP*) errors, and the total misclassifications cost calculated in the confusion matrix (Ting 2002, Maimon and Rokach 2005). As an effect of the initial weight setting in the cost matrix, where the false positives (*FP*s) have been weighted higher than the false negatives (*FN*s), the confusion matrix will have fewer *FP*s than *FN*s (Witten and Frank 2005). All the setting values used to construct 40 models, when the cost matrix optimization was applied, are displayed in Appendix D (Table D1), while the settings of the models with the greatest accuracy are shown in Table 5.2.

5.2.4 Modelling and Analysis

5

First, the accuracy of the DTMs was evaluated with two measurements obtained from the confusion matrix: the number of correctly classified instances (CCI) and Cohen's Kappa statistic (κ) (cf. Section 2.3.2).

When the cross-validation results, which are calculated beginning with the confusion matrix, are slightly different, it is difficult to determine in the first instance which measurement is better for evaluating a decision tree model (DTM). Furthermore, the accuracy of the DTM (i.e., CCI) is uniquely obtained regardless of how the other off-diagonal elements take their values (Wei et al. 2010). The misclassification information (i.e., FP + FN) of confusion matrices can be analyzed using the measurement of the overall confusion entropy of a confusion matrix (CEN). According to Wei et al. (2010), higher accuracy of the models is likely to correspond to lower confusion entropy. Likewise, the CEN is more precise than the correctly classified instances (CCI), and can replace this latter coefficient to evaluate classifiers in classification applications. In addition to the CCI, the least confusion entropy was considered as a decision value to choose the best model for each analyzed regulation. For this calculation, the following expression was adapted to a confusion matrix of 2 × 2 (Eq. 5.1), from equations given by Wei et al. (2010):

$$CEN = (P_1 + P_2)CEN_j \tag{5.1}$$

where, P_j is called confusion probability of class j and CEN_j is defined as confusion entropy of class j. These values were calculated with the next expressions Equations (5.2) and (5.3).

$$P_1 = \frac{TP + FN}{2(TP + FN + FP + TN)} \; and \; P_2 = \frac{FP + TN}{2(TP + FN + FP + TN)} \tag{5.2}$$

$$CEN_j = -P_{FN} log_2 P_{FN} - P_{FP} log_2 P_{FP} \tag{5.3}$$

In Equation (5.3) P_{FP} and P_{FN} are the misclassification probability of classifying the samples of class i to class j subject to class j, are defined in Equation (5.4).

$$P_{FP} = \frac{FP}{FN + FP + 2TP} \; and \; P_{FN} = \frac{FN}{FN + FP + 2TN} \tag{5.4}$$

In order to check the stability of the DTMs, and knowing that the dataset is relatively small, the dataset was randomly and manually divided into three subsets, and stratified based on fulfillment or non-fulfillment of the regulation in analysis. Two of these subsets were used to train the model, and the third subset was applied to test the model. This process was repeated three times so that each subset was used to check the others. The Stability of the DTMs was calculated from the wide variation of the standard deviation (Goethals 2005) of the CCI and Kappa statistics (Forio et al. 2016) that were obtained from the test subsets.

5

The model optimization, in addition to the statistical fit, must be assessed from an ecological point of view. In some cases, erroneous results from an ecological angle could also occur. For this reason, before choosing a model, an ecological examination has to be considered (Everaert et al. 2013b), in which the obtained rules from the DTM are compared and tested for what is generally accepted in ecology (Everaert et al. 2013a). Thus, in this research for the ecological evaluation, two criteria were included. The first, the DTM, was discarded when a taxon resulting from the model had a tolerant score (TS) lower than four, which ensured that the microbial water quality assessment was not done in a highly polluted place. The second criterion was to ensure that at least one of the taxon resulting from the DTM was always present. In some cases, it is possible to obtain from the branches (rules) of a DTM that the presence of any taxon is not necessary for compliance with the fecal coliform regulation. This is an aspect that could give erroneous results on the application of the DTM.

Finally, the selected models, after optimization process were assessed with two new datasets taken both in dry (July of 2015) as well as in rainy (March of 2016) seasons.

5.3 Results

5.3.1 Current Water Quality Status

Fecal coliforms (FCs) concentrations were greater in urban and suburban sites than sites from other land uses (Fig. 5.1). According to the 25% and 75% percentiles, in urban areas, the variation of FCs was lower than suburban areas (Fig. 5.1). Similarly, the FCs were higher in pasture areas than native vegetation areas, with higher values in the 25% and 75% percentiles. The variation of FCs was the lowest in forest areas (Fig. 5.1). In addition, a summary of the variation of the physicochemical parameters collected during the sampling campaign can be reviewed in Table 2.2.

Similarly, the fecal coliforms results, in relation to the three microbial water quality standards described in Section 5.2.1 and with land use (Fig. 2.2A, Table A5), show that the nine points sampled in the south–east section of the basin that are located in the urban and suburban areas of Cuenca, do not meet the official microbial water quality standards (Fig. 5.2A–C). All other locations (24 points) meet the regulation standards regarding agriculture (<1000 MPN.100 mL^{-1}) and raw water (<2000 MPN.100 mL^{-1}). It is important to note that these 24 sites met both regulations at the same time. Additionally, nine points previously indicated, five sites are not meeting the recreational regulation (<200 MPN.100 mL^{-1}). The location of the aforementioned five sites is close to livestock zones: three are near the center of the Machangara basin (points: 24, 27 and 45), and two other locations are in the northeast area of this catchment (points: 13 and 15) (Fig. 5.2A).

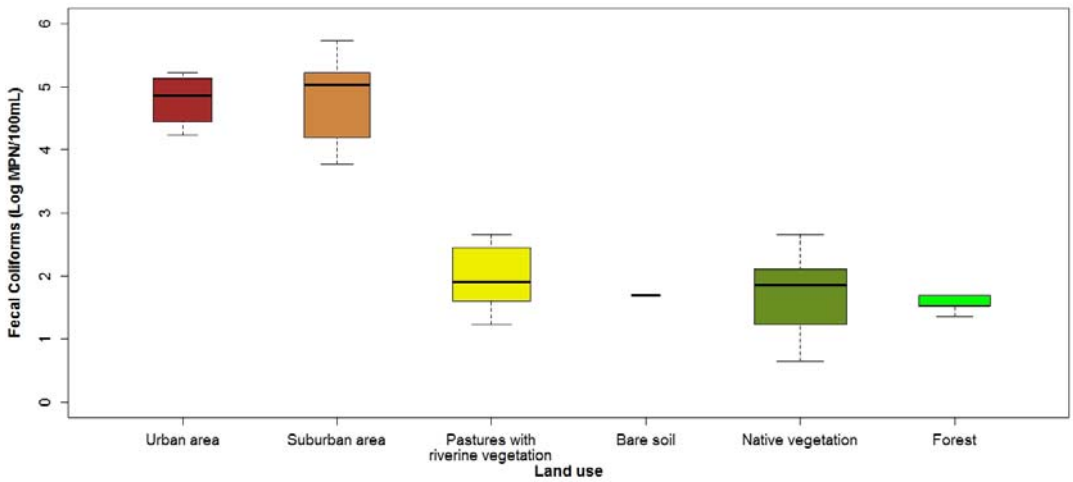


Fig. 5.1. Boxplots of the fecal coliforms variation according to land use.

When analyzing the results of the BMWP-Col of the 33 points indicated in Fig. 2.2B, in relation to the fecal coliforms regulations, the outcomes show 14 points with different biological water quality (i.e., two good, eight moderate, two deficient, and two bad) that do not meet the recreational fecal regulation. In addition, nine points with diverse BMWP-Col (i.e., five moderate, two deficient and two bad), do not meet the values of the agriculture and raw water fecal regulations.

5.3.2 Model Development

For the construction of the models, 23 taxa that were observed in at least three different points were used (Table A4 and Table A5). In total eight models were constructed during the model development stage, from which four resulted from the absence-presence dataset while that four other models developed with the abundance dataset (Table 5.2 and Table D2 in Appendix D). Based on the correctly classified instances (CCI) and Kappa statistics, a reliable decision tree model (DTM) (i.e., CCI > 70% and k > 0.4) obtained from models 2a1 and 2a2 (Table 5.2 and Table D2 in Appendix D), was developed with the abundance database, allowing a preliminary assessment of the fulfillment of the agriculture fecal coliform guidelines (Fig. 5.3B). No reliable model was obtained to assess the fulfillment of the recreational fecal coliform regulation. Similarly, from the presence-absence database no confident DTMs (Table

5.2) were obtained to check the accomplishment of any fecal coliforms guidelines. With the dataset used, it was not possible to obtain a specific model to verify raw water regulation, although the model obtained for agriculture fecal regulation, which has a more stringent threshold, could be adopted to check the raw water regulation. Likewise, the two best models had as a result, the same DTM (Models 2a1 and 2a2—Table 5.2), whose description is shown in Section 5.3.3 following an ecological examination.

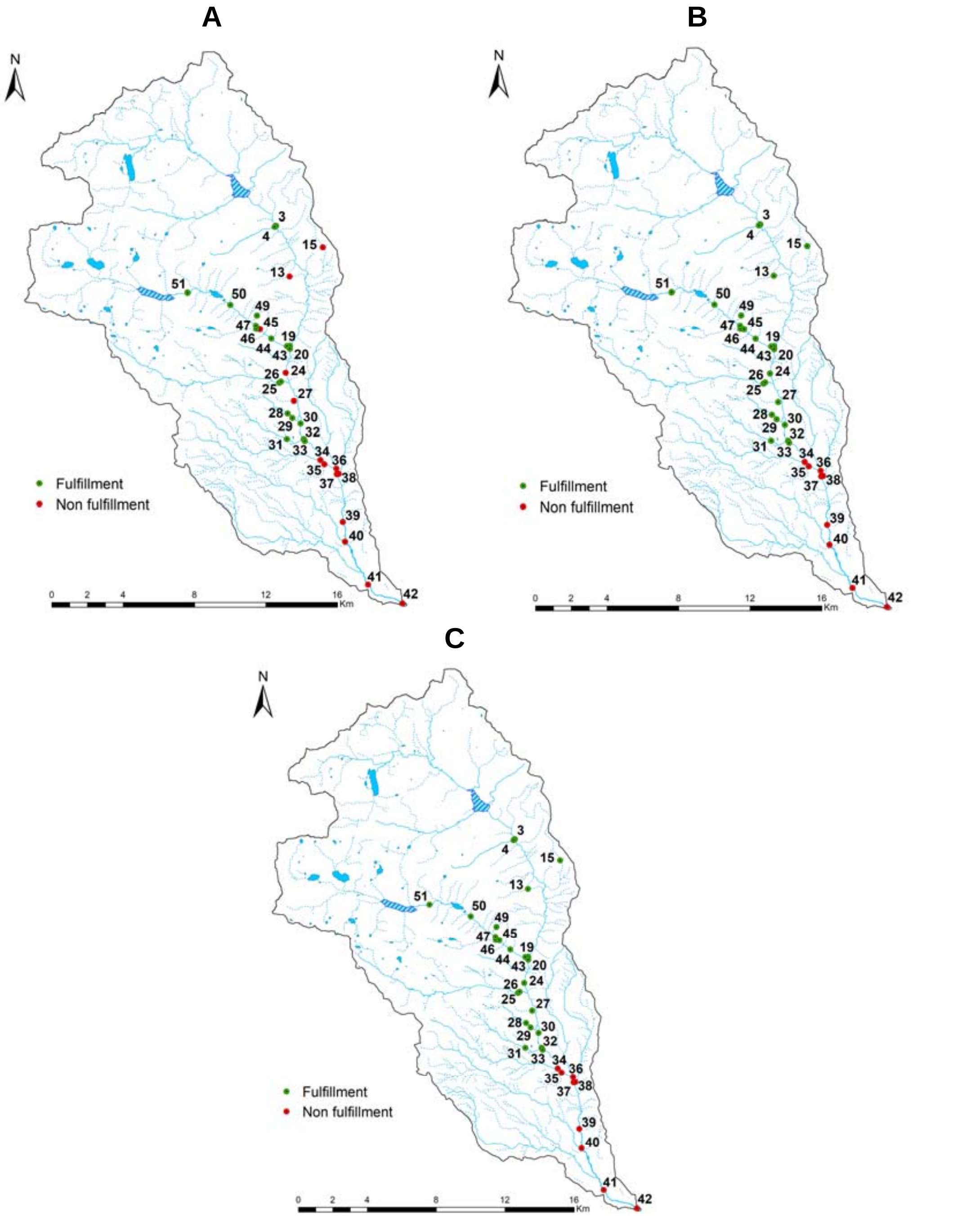


Fig. 5.2. Fulfillment of fecal coliforms limits in relation to water use. (**A**) Recreational with primary contact; (**B**) Agricultural and livestock use; (**C**) For raw water use previous to non-conventional treatment required.

5.3.3 Model Optimization

The decision tree models (DTMs) were optimized adding new weights to false positives in training instances, with the aim to minimize the false positive (*FP*) errors. This is possible with a cost-sensitive classifier (CSC) tool with the J.48 algorithm in the WEKA package. It was not possible to obtain a specific model for the raw water fecal regulation, but the resulting DTMs obtained from the agriculture regulation could be applied to check the raw water fecal regulation. Moreover, the threshold of the agriculture regulation is more stringent than the raw water fecal coliform regulation. In this stage 40 models were developed (Table D3 in Appendix D), from which eight DTMs were reliable (Table 5.3), with their correctly classified instances (CCI) higher than 0.7 and with their Kappa statistics higher than 0.4 (Table 5.3). These eight DTMs were initially pre-selected from a statistical point of view (Table 5.3). Two groups of models for evaluation of the recreational fecal coliform regulation had similar trees with different abundance requirement (models from 1a5 to 1a7 and from 1a9 to 1a12), that group which had the model with the least entropy of a confusion matrix (CEN) was chosen (models from 1a5 to 1a7). For the agriculture fecal regulation, the DTM resulting from models 2a3 to 2a6 was the same that was obtained in models 2a1 and 2a2 in the previous section, “Model development”. The DTMs achieved from models 2a3 to 2a6 and from 2a7 to 2a11 had the same families with the same requirements of abundance, differing between both DTMs the sequence of their leaves. In this case, the group that had the model with the least CEN was selected (models from 2a3 to 2a6), resulting in six total DTMs after statistical evaluation (1a4, 1a5 to 1a7, 1a8, 2ap3, 2ap4 and 2ap5, and 2a3 to 2a6—Table 5.3). All models that were obtained after the optimization process with their results of the correctly classified instances (CCI), Kappa statistics, the number of leaves obtained in each model through k-fold (i.e., three, five and 10) cross-validation and the overall confusion entropy of a confusion matrix (CEN), are shown in Table D2.

Table 5.3. Summary of the predictive result of the models with the best accuracy after optimization, in which the cost matrix weights was used: correctly classified instances (CCI), Kappa statistics and overall confusion entropy of a confusion matrix (CEN).

Model No.	Model Settings						Model Outcomes		
	J4.8	PCF [a]	CMW [b]				CCI [g] (%)	Kappa Statistics	CEN [h]
			TP [c]	FN [d]	FP [e]	TN [f]	CCI ± sd	Kappa ± sd	CEN ± sd
1[i]a[j]-4[k]	3, 5 and 10 fcv [l]	0.25	0	1	2	0	72.7 ± 6.1	0.4 ± 0.1	0.8 ± 0.1
1a5 to 1a7	**3, 5 and 10 fcv**	**0.25**	**0**	**1**	**3 to 5**	**0**	**77.4 ± 8.0**	**0.6 ± 0.2**	**0.6 ± 0.1**
1a8	3, 5 and 10 fcv	0.25	0	1	7	0	78.8 ± 8.0	0.6 ± 0.2	0.6 ± 0.1
1a9 to 1a12	3, 5 and 10 fcv	0.1 and 0.25	0	1	8 and 9	0	74.9 ± 6.9	0.5 ± 0.1	0.7 ± 0.1
2ap3	3, 5 and 10 fcv	0.25	0	1	2	0	75.8 ± 5.3	0.5 ± 0.1	0.7 ± 0.2
2ap4 and 2ap5	3, 5 and 10 fcv	0.25	0	1	3 and 5	0	72.7 ± 7.2	0.4 ± 0.1	0.7 ± 0.1
2a3 to 2a6	**3, 5 and 10 fcv**	**0.25**	**0**	**1**	**1 to 4**	**0**	**87.1 ± 6.6**	**0.7 ± 0.2**	**0.5 ± 0.2**
2a7 to 2a11	3, 5 and 10 fcv	0.25	0	1	5 to 15	0	80.2 ± 9.4	0.6 ± 0.2	0.6 ± 0.1

Mean and standard deviations of CCI, Kappa statistics and CEN were derived from k-fold cross-validation. [a] PCF = Pruning confidence factor; [b] CMW = Cost Matrix Weights; [c] TP = True positives; [d] FN = False negative; [e] FP = False positive; [f] TN = True negative; [g] CCI = Correctly classified instances; [h] CEN = Overall confusion entropy of a confusion matrix; [i] Fecal coliform regulation: 1 for recreational and 2 for agriculture; [j] Kind of database: ap = absence/presence, a = abundance; [k] Number of model with different value of PCF; [l] fcv = folds cross-validation. The chosen models are highlighted in bold.

These pre-selected decision tree models (DTMs) were verified from an ecological point of view. Three group of models were discarded: the first with the model 1a4, the second with the model 2ap3, and the third with the models 2ap4 and 2ap5 (Table 5.3 and Table D3 in Appendix D). *Chironomidae*, which is a taxon with very low pollution sensitivity, is present in the leaves of the first discarded 1a4 DTM (Table 5.3). This 1a4 model was constructed with the abundance dataset to assess the fulfillment of the recreational fecal coliform regulation, while, the second (model 2ap3) and third (models 2ap4 and 2ap5) DTMs were developed with the absence-presence dataset, to evaluate the accomplishment of the agriculture fecal coliform regulation. In those DTMs, the rules are determined by the presence and absence of *Perlidae* and *Baetidae* taxa. The absence of both aforementioned sensitive taxa meets the agriculture fecal coliform regulations (Table D5 and Table D7 in Appendix D). However, this situation can also register in polluted sites. The three remaining DTMs (1a5 to 1a7, 1a8 and 2a3 to 2a6 — Table 5.3) were evaluated with the validation datasets, from which two models were confirmed (1a5 to 1a7 and 2a3 to 2a6—Table 5.3), whereas the one DTM obtained from model 1a8 (Table 5.3), constructed for verification of recreational water use with primary contact guidance, could not be validated nor discarded. Validation was not possible due to the fact that the latter DTM did not meet with the requirement of abundance given by its second branch (Table D6 in Appendix D). This, despite the fact that the first branch of the model met the FC regulation and was validated. Finally, two decision tree models (DTMs) were selected (from models 1a5 to 1a7 and from models 2a3 to 2a6 — Table 5.3), in which the abundance of each taxon refers to the number of specimens collected in five square meters (5 m^2). The first DTMs is applicable as preliminary tools for verification of recreational water use with primary contact guidance, which is referred to in this work as the 'recreational fecal regulation'. This first DTM (from models 1a5 to 1a7 — Table 5.3) has as a condition, the presence of *Baetidae* (Ephemeroptera) with an abundance less or equal to three and the presence of *Scirtidae* (Coleoptera) with an abundance minor or equal to four (Fig. 5.3A). The second DTM (from models 2a3 to 2a6 — Table 5.3) is used as a proxy indicator to evaluate the success of the agriculture fecal standards that regulate agriculture and livestock water uses. This second DTM (from models 2a3 to 2a6 — Table 5.3 — Fig. 5.3B) was the same that was obtained before the optimization step (models 2a1 and 2a2—Table 5.2). The model showed that the presence of *Perlidae* (Plecoptera) is necessary, if this taxon is not present, *Baetidae* (Ephemeroptera) must have an abundance of one but less than or equal to four. If its abundance is higher, the non-fulfillment of the regulation is complete. The rules generated by the leaves of the chosen DTMs were also checked with the fulfillment of the recreational and agriculture fecal coliforms regulations (Table D4 and Table D5 in Appendix D), as well as the validation datasets (Table D6 and Table D7 in Appendix

D), verifying that all points that met the requirements of the DTMs satisfied the analyzed fecal coliforms standards.

The stability of the models of the same class (e.g., 3-fcv and 0.10 as PCF) was determined by the variation among correctly classified instances (CCI) and Kappa statistics obtained from the tree fold cross-validation. The results shown in Appendix D (Table D8 and Table D9), demonstrate that on average the standard deviation represents 20% of the mean of the CCI and 61% of the mean of the Cohen's Kappa statistics, for the models of the recreational fecal regulation. While, for the agriculture regulation models, the standard deviation is, on average 14% of the CCI and 73 % of the Kappa statistics. This revealed that the CCI deviation was acceptable, while for the Kappa statistics the variation range was high.

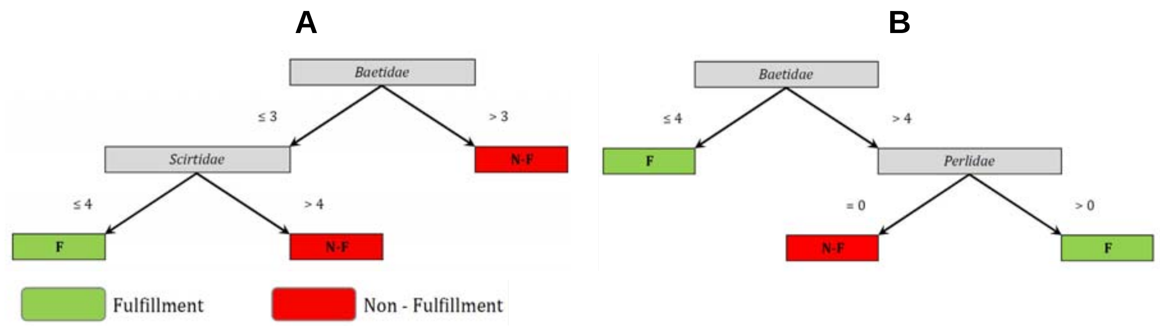


Fig. 5.3. Macroinvertebrates abundance decision tree models (DTMs) in relation to fecal coliforms water use standard. Fulfillment of: (**A**) primary contact and (**B**) agriculture and livestock irrigation.

5.4 Discussion

5.4.1 Model Relevance and Optimization from a Statistical Point of View

Classification trees successfully modeled the abundance of some macroinvertebrates taxa as a proxy indicator of the fulfillment of two Ecuadorian fecal coliform regulations for water use. One decision tree model (DTM) was obtained in the development stage, and one after the optimization phase. Furthermore, both DTMs were also confirmed with the validation datasets. In both cases, the models had a maximum of three variables that were hierarchically structured as levels of knowledge, allowing their rules to be easily applied (Hoang et al. 2010b). However, the inclusion of a large number of variables would result in a complex DTM with many rules that would hamper its application (Džeroski et al. 1997). Additionally, this technique is non-parametric and non-linear. Consequently, the explanatory and response variables are not assumed to have a linear relationship (Moisen 2008).

It was not possible to obtain a specific DTM to check the raw water coliform regulations. This was because the same locations satisfied both agriculture and raw water regulations. However, the DTM obtained to verify the agriculture regulation could be used to check the raw water fecal coliform regulation, as the threshold of the agriculture regulation is more stringent. With a new dataset, in which the occurrence of sites that meet only the raw water coliform regulation, a specific model for checking the fulfillment of this standard could be constructed. Before the optimization phase, no models were obtained with the presence-absence dataset, while a DTM was only found with the abundance dataset. Most likely, it happened because the presence-absence dataset was binary (i.e., 0 and 1), while with the abundance dataset, the classification tree technique probably had more attributes to construct the rules of classification. Thus, Maimon and Rokach (2005) noted that with the use of binary data the manipulation of categorical data is simplified and its normalization is eliminated, which makes it more difficult for binary data to be clustered. From a statistical point of view, after the optimization process in which the false positives errors were more costly than the false negatives (Witten and Frank 2005), two DTMs (model 2ap3 and from models 2ap3 and 2ap4 — Table 5.3) were obtained from the presence-absence information, and six models resulted from the abundance datasets (1a4, 1a5 to 1a7, 1a8, 1a9 to 1a12, 2a3 to 2a6 and from models 2a7 to 2a11 — Table 5.3). For the recreational fecal coliform regulation, it was not possible to construct a reliable model with the presence-absence dataset. While with the abundance dataset, the same decision tree model for the agriculture fecal coliform regulations was achieved before and after the optimization process until the false positives (*FP*s) were weighted four times, with the help of a cost-sensitive classifier (CSC). When the *FP* was weighted from five to 12, the rules generated by the trees changed their order, resulting in a new, reliable DTM (from models 2a7 to 2a11 — Table D1 – Table D3) with the same final outcomes as the previous DTM (from models 2a3 to 2a6 — Table D1 – Table D3). The maximum correctly classified instances (CCI) and Kappa statistics and the least confusion entropy of a confusion matrix (CEN) were obtained when the *FP* was weighted twice, yet with higher weighted values than 12, unreliable decision tree models were obtained. The DTMs resulting from the abundance dataset for the recreational fecal regulations (models: 1a4, 1a5 to 1a7, 1a8, and from 1a9 to 1a12—Table D1 and Table D2), were shown to be reliable when the *FP* was weighted from two to nine with the CSC, arriving at the maximum CCI and Kappa statistics and the least CEN when the weighted value was seven (model 1a8 — Table D1 and Table D2). In this regard, Maimon and Rokach (2005) showed that to select the optimum value of weighted false positive requires a sensitive analysis of the effect of its value on the accuracy of the resulting model. During the optimization process, two groups of reliable decision tree models (DTMs) (models from 1a5 to 1a7 and from 1a9 to 1a12—Table 5.3), constructed to evaluate the recreational fecal regulation, showed the same trees with the same taxa, but with different abundance requirements. When the abundance was higher, the DTM was more reliable. This was likely due to the WEKA trying to increase the model accuracy when the false positives (*FP*s) were reweighted with the cost-sensitive classifier (CSC), an increase of *Scirtidae* was required since the size of the dataset was relatively small.

With regard to the stability of the models, classification trees with relatively small datasets tended to be unstable (Goethals 2005), a pattern that also was found in the selected DTMs. Thus, with the analysis of the variation of the correctly classified instances (CCI) and Kappa statistics, the first parameter (i.e., CCI) appeared more stable than the Kappa statistics. This typically happens when a dataset is relatively small, and each database has limited extractable information, so accordingly, Kappa statistics values represent the information content of the dataset (Ambelu et al. 2010).

5.4.2 Model Relevance and Optimization from an Ecological Point of View

With regard to the organic pollution tolerance of taxa, the BMWP-Col index gives a sensibility score range with one being the most tolerant families, to 10 being the less tolerant macroinvertebrates. Thus, the tolerance values of the taxa shown in the final decision tree models (DTMs) shown in Fig. 5.3 were 10 for *Perlidae* (Plecoptera), six for *Scirtidae* (Coleoptera), and five for *Baetidae* (Ephemeroptera) (Roldán Pérez 1999). In the Ecuadorian Andes, *Scirtidae* was found in clean and slightly polluted rivers, while *Baetidae* was found in places that were clean as well as in some polluted sites, but not in very polluted points (Jacobsen 1998). In the selected two DTMs, an abundance higher than four of these taxa was a negative indicator of the water quality. Likewise, *Perlidae* was present in pristine conditions and unpolluted places in the Andes of Ecuador (Jacobsen 1998, Ríos-Touma et al. 2011). With regard to the relationship between fecal coliforms and biological water quality in the Cuenca River basin, it was found that fecal coliforms were the explanatory variable for the presence of *Physidae (Holguin-Gonzalez et al. 2013a),* which has a low tolerance score of three (Roldán Pérez 1999), in places where the biological water quality varied from poor to moderate. In the same way, it was established that one of the explanatory variables for the *Perlidae* presence was fecal coliforms (Jerves-Cobo et al. 2017). Whereas, Acosta and Hampel (2015) in the Cuenca River basin, found that fecal coliforms were unique variables that had relative importance in the distribution of the macroinvertebrate communities in the rivers of the moorland. These authors also pointed out that fecal coliforms influenced the structure of the benthic communities in rivers with urban influence. The two final DTMs (models from 1a5 to 1a7 and from 2a3 to 2a6 — Table 5.3 and Fig. 5.3), chosen in this research, show that the three taxa of macroinvertebrates (i.e., *Perlidae, Scirtidae* and *Baetidae*), may be sensitive to fecal pollution.

The decision tree models (DTMs) resulting from model 1a4 (Table 5.3 and Table D3), which constructed for recreational regulation analysis with the abundance dataset, did not pass the ecological examination due to the presence of *Chironomidae,* whose tolerance score is two, was present in one of its leaves (rules). This situation could have been due to the identification of *Chironomidae* that was analyzed to family level and not to a sub-taxa level. In some instances, this kind of identification such as the subfamilies of *Chironomidae* includes species with large differences in tolerance to pollutants (Džeroski et al. 2000). Similarly, two DTMs obtained from model 2ap3 and from models 2ap4 and 2ap5, which were constructed with the presence-absence dataset, were discarded. In both DTMs, the agriculture regulation was accomplished without the presence of *Perlidae* and *Baetidae.* However, the presence of both aforementioned taxa was not registered in polluted and very polluted places (Jacobsen 1998, Ríos-Touma et al. 2011), that could give both DTMs erroneous outcomes; although, both models could be modified, retaining only the part of the decision trees that could give reliable results. In this regard, in data mining models such as decision trees, a single model can be modified into multiple models, and the resulting models can operate in a large variety of conditions (Maimon and Rokach 2005).

The percent of occurrence of the analyzed taxa in the sampled sites in the Machangara River basin was as follows: *Perlidae* 36%, *Scirtidae* 36% and *Baetidae* 82%. The absence of these taxon in other areas may be due to specific reasons. For example: the habitat in some places may be unstable (Jacobsen 2005), or it was not suitable for a specific taxon (Burneo and Gunkel 2003, Dallas and Day 2007). While in suitable environments, a taxa could be temporarily absent, for example, due to migration or seasonal variation (Jacobsen 1998). Likewise, the abundance, a fundamental parameter in the chosen decision tree models (DTMs), may also fluctuate seasonally. Thus, Jacobsen (1998) found that the density of macroinvertebrates is much higher in the dry season than in the rainy season in the Ecuadorian highland streams. Although a higher abundance of macroinvertebrates would not influence the final results of the chosen DTMs as the maximum threshold of the required abundance is four. In some areas that were close to livestock, the concentration of fecal coliforms in the river was low. Perhaps the riparian habitats were able to uptake pollution transported by run-off from livestock areas (Kauffman and Krueger 1984), or the run-off volumes were small and only transported minimal amounts of pollutants into the streams and rivers. Or also, the river places located below livestock areas experienced an unstable habitat from recurring shifts in pollutants concentration, especially during rainy season.

5.4.3 A Possible Screening Tool for Microbial Pollution

The intake of microbiologically contaminated water is a great concern from a human health perspective (Fewtrell et al. 2001). The main sources of organic pollution to surface water are: wastewater, storm water outfalls, as well as livestock and wildlife feces (Oliver 1984, Seyfried and Harris 1990). Additionally, the presence of pathogens in the water shows a good correlation to the presence of fecal contamination (Leclerc et al. 2001). As a result, fecal bacteria or thermal-tolerant bacteria have been used as the main indicators of fecal pollution and also the possible presence of disease-causing organisms (Fewtrell et al. 2001, WHO 2003, Tallon et al. 2005).

The procedure to sample and to identify the presence or absence and abundance of selected macroinvertebrates families in a river takes less than one hour, by individuals who have been trained in identification and sampling protocols. This activity can applied in the field by a person with minimal training. Since the models allow personnel to focus on a few taxa of key indicator importance, not all taxa need to be identified, and the focus can be placed on searching for particular groups. Conversely, standard methods to measure fecal indicator bacteria for recreational, irrigation or drinking water uses require at least 24 h to obtain results. For the detection of *E. coli* or thermo-tolerant coliforms, several methods have been recommended by the International Organization for Standardization (ISO), including procedures such as most probable number (MPN) (Oliver 1984). Furthermore, this detection of water pollution needs to be performed at least daily (Wade et al. 2006). In contrast, the models (DTMs) introduced in this work could be used as inexpensive (proxy) bioindicators for fecal contamination that do not require laboratory support or highly qualified personnel. As a result of this research, the application of the decision tree models (DTM) is a simpler and faster method as a proxy indicator to assess fecal pollution in rivers.

It is important to note, that the two decision tree models (DTMs) introduced and chosen for application in this research, can be improved both by collecting more data from the same sites in different seasons and by collecting more data from new sites in the Machangara River basin in the dry and rainy seasons. Thus, the taxa variation between two seasons (Jacobsen 1998) can be included. This new data can be used to update the current DTMs. The models introduced in this work should also be tested in different river basins before being applied in other locations, due to the variation of environmental conditions such as weather, vegetation, and soil use. For this reason, it is recommended that samples be taken from different locations in relation to land use. According to Forio et al. (2015), testing these models in a wider range of situations over time, will permit researchers to define the range of applications for which the model predictions are suitable. Additionally, after their first application, the results must be confirmed in a laboratory using traditional analysis.

5.5 Conclusions

Decision tree models (DTMs) were developed as preliminary assessment tools to check the compliance to two Ecuadorian microbial water quality standards associated with fecal coliforms. These DTMs were based on the presence and abundance of *Perlidae, Scirtidae* and *Baetidae* in the Machangara River basin located in the southern Andes Mountains of Ecuador. These DTMs showed that an abundance higher than four of the aforementioned taxa, with the exception of *Perlidae,* is an indicator of fecal coliforms pollution. The two best-performing models were adopted and can be applied by personnel with minimum training in the identification of the aforementioned taxa. The use of the cost-sensitive classifier (CSC) in the Waikato Environment for Knowledge Analysis (Weka) package to eliminate false positives (*FP*) in the confusion matrix improved the reliability of the resulting models. The models introduced in this work still need to be tested over time to ensure their stability (and reliability), before being applicable to areas with sources of fecal pollution. It needs to be stressed that these tools will not eliminate microbial tests, but can serve as a rapid screening process and moreover, allow the detection of key indicator invertebrate taxa related to water quality.

Chapter 6. Biological water quality in tropical rivers during dry and rainy seasons: A model-based analysis

Adapted from:

Rubén Jerves-Cobo, Marie Anne Eurie Forio, Koen Lock, Jana Van Butsel, Guillermina Pauta, Félipe Cisneros, Ingmar Nopens and Peter L. M. Goethals. (submitted). Biological water quality in tropical rivers during dry and rainy seasons: A model-based analysis.

Abstract

Biological studies indicate substantial differences between seasons in tropical rivers. However, investigations on the needs and added value of season-specific models are lacking. This chapter aims to determine the accuracy and relevance of season-specific and season-overarching models to predict biological water quality. Additionally, the variation of prediction accuracy using sub-datasets from different parts of the Cuenca River basin was investigated. 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 the five-day biological oxygen demand (BOD_5), ammonium and orthophosphate, while dissolved oxygen (DO), oxygen saturation (OS), nitrate and total solids proved to be important during the rainy season. The results of this chapter 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 modelling approach developed in this chapter can be applied to similar basins in the tropics.

6.1 Introduction

Efforts to restore the good ecological status of water bodies throughout the world have been conducted in urban and suburban areas, 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 other discharges 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. 2009b). 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. 2014). Hence, restoration options must be evaluated using scenario analyses before their application in order 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 (Ambelu et al. 2010, Everaert et al. 2014). 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). Thus, there is a need to select the most appropriate models. These models have been recognized as powerful tools of predictive ecology (Holguin-Gonzalez et al. 2013b) and are instruments that can be used in the restoration and conservation of aquatic ecosystems (Mouton et al. 2009b). 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, Hoang et al. 2010a, 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 over the other 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 with 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 establishing the main variables that contribute to obtaining accurate predictions using small datasets (Vayssières et al. 2000).

The framework and objectives of the current chapter are presented in Fig. 6.1. The aim of this chapter is to develop 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. These ecological models are called ecological assessment models (EAM). Furthermore, the predictability of sites throughout seasons and between the dry and the rainy season based on simulations was analyzed. An analysis was also included of 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 findings of this chapter can be applied to scenario analyses for river restoration based on the key disturbance variables obtained from the models. As indicated in the framework, the cause deduction (phase two in Fig. 6.1) is not addressed in this chapter because this requires additional analysis and simulations, either in the actual rivers or in river water quality models.

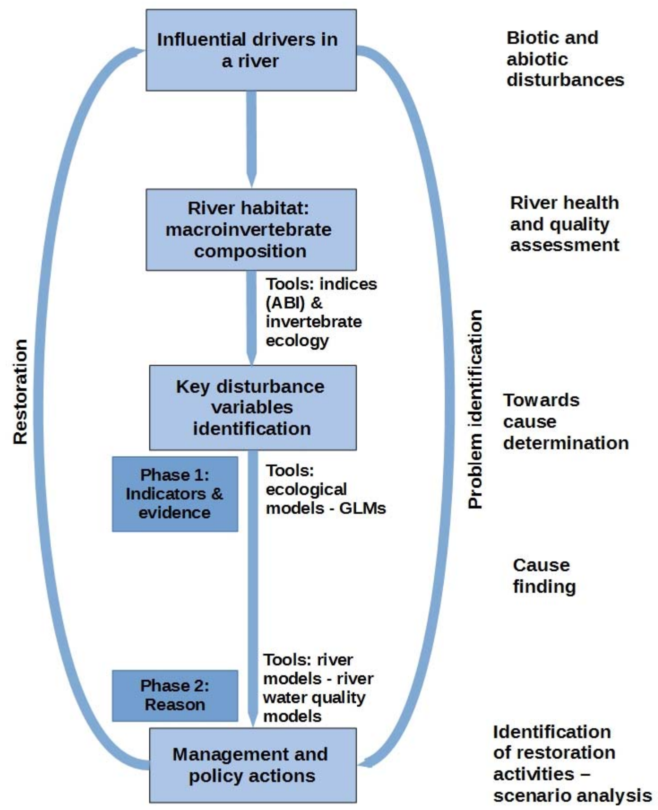


Fig. 6.1. The applied framework for the restoration of river water quality, of which Phase One was implemented in this chapter (adapted from Forio et al. (2017)).

6.2 Data analysis

6.2.1 Model development and validation

Ecological models were developed to determine parameters, which could influence the performance of the Andean Biotic Index (ABI) in the study area. For this, the data was fitted by a generalized linear model (GLM).

To build the models, five datasets as described in Fig. 6. 2A were fitted with Gaussian, Gamma and Inverse Gaussian GLMs (cf. Section 2.3.3). 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 according to calculations and plots from the Cook's distance and the hat elements (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 comprised the following: chemical oxygen demand (COD), chlorophyll-a, iron, nickel, tannins + lignins, copper, aluminum, silicon, chrome, manganese and fluorides. A total of 18 variables were considered during model development of each dataset. These variables were 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 (TSol), turbidity, flow velocity, water temperature, pH, conductivity, alkalinity, true color and bank material. The bank material surveyed (Jerves-Cobo et al. 2018b) 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 ±0.6. In this regard, Zuur et al. (2009) indicated a value of ±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 (cf. Section 2.3.3). Subsequently, two sub-datasets were used for training the model, while the third was used for testing (validation) (Fig. 6. 2A 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.

6.2.2 Variable selection

The selection of explanatory variables that have an influence on the biological index ABI, a stepwise backwards selection procedure was implemented (cf. Section 2.3.1). These stepwise 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* less than 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 less than 0.10. All statistical tests were performed at the 5% significance level.

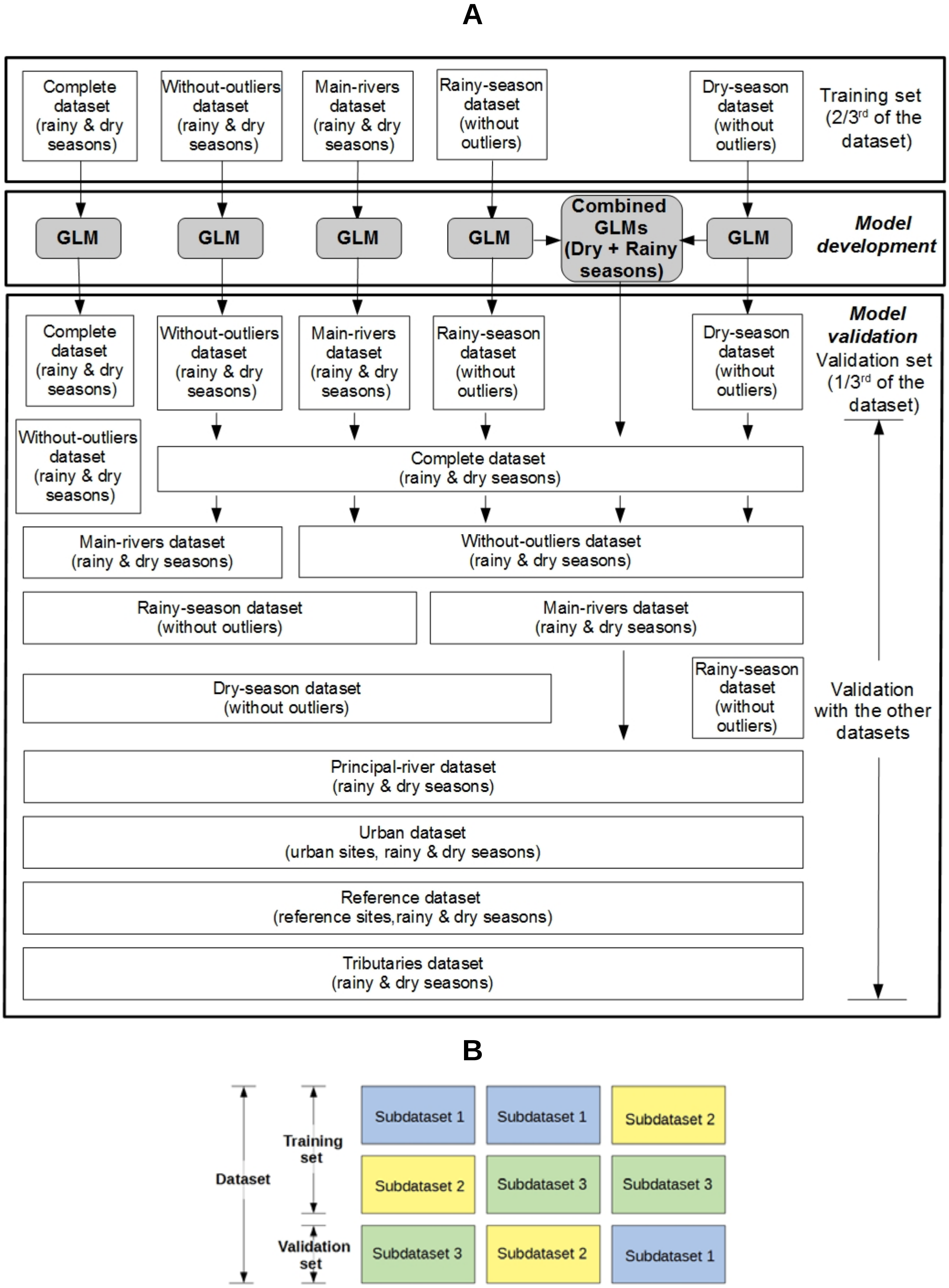


Fig. 6. 2. Model development and validation scheme: (A) General schema of the models construction and (B) three-fold cross-validation procedure. The GLM represents generalized linear models.

6.2.3 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. the third sub-dataset). Similarly, the goodness-of-fit of each model 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 2.2.3. The accuracy of the models was evaluated with the correctly classified instances (CCI) and Cohen's Kappa statistic (κ). The criteria of the model selection was described in section 2.3.2.

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. 6. 2A). In addition, these models were validated with sub-datasets 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. 6. 2A). 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. 6. 2A). The sites included in each dataset are presented in Table 6.1 and Fig. B7.

6.3 Results

6.3.1 Variable differences between seasons and influential points analysis

According to the results of the paired Student *t*-test and paired Wilcoxon test, when the same sampling sites between both seasons were compared (Table E1), they indicated that the total number of taxa, the Shannon-Wiener index (SWD), water temperature, pH, DO, total solids and Nitrite were significantly different between rainy and dry seasons. However, turbidity, oxygen saturation, alkalinity, chloride, ammonium and flow velocity did not demonstrate 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. E1A to Fig. E1D). However, no influential point was detected for Gaussian distribution (Fig. E1E and Fig. E1F). Despite this, the aforementioned site was excluded from the other datasets. 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 (Fig. E2 to Fig. E5). Furthermore, the site Ta01 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. E3).

Table 6.1. 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		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	

6.3.2 Season-specific models

In this section and forward, only the models that were performing the best, that is, models that obtained the highest κ and CCI values during the validation process were presented. 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 (Table 6.2 and Table E2). However, ammonium and chloride were also found as predictors of the ABI in one of the four chosen models (Table E2). Gamma and Inverse Gaussian GLMs predicted the ABI better than the Guassian GLM. However, results suggest that Gamma was the best performing GLM.

The models obtained to predict the ABI during the rainy season (Table 6.2 and Table E3) 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 E3). Nitrite and bank material were 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 E3). Fig. 6.3 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. E6.

Table 6.2. The best-performing season-specific models.

		Season-specific models					
Explanatory Variables	Regression Parameters	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	< 0.01	-725.5	195.9	< 0.01
Nitrate	B1				85.3	26.6	< 0.01
Nitrite	B2				-1100.7	319.9	< 0.01
Ammonium	B3	-4.2E-02	1.7E-02	0.03			
BOD_5	B4	5.2E-03	1.3E-03	< 0.01			
DO	B5				-42.8	19.7	0.04
Oxygen Saturation	B6				10.9	2.8	< 0.01
Orthophosphate	B8	1.4E-01	4.7E-02	< 0.01			
Total solids	B10				-0.1	0.1	0.03
Bank material	B18				10.6	4.1	0.02
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**	

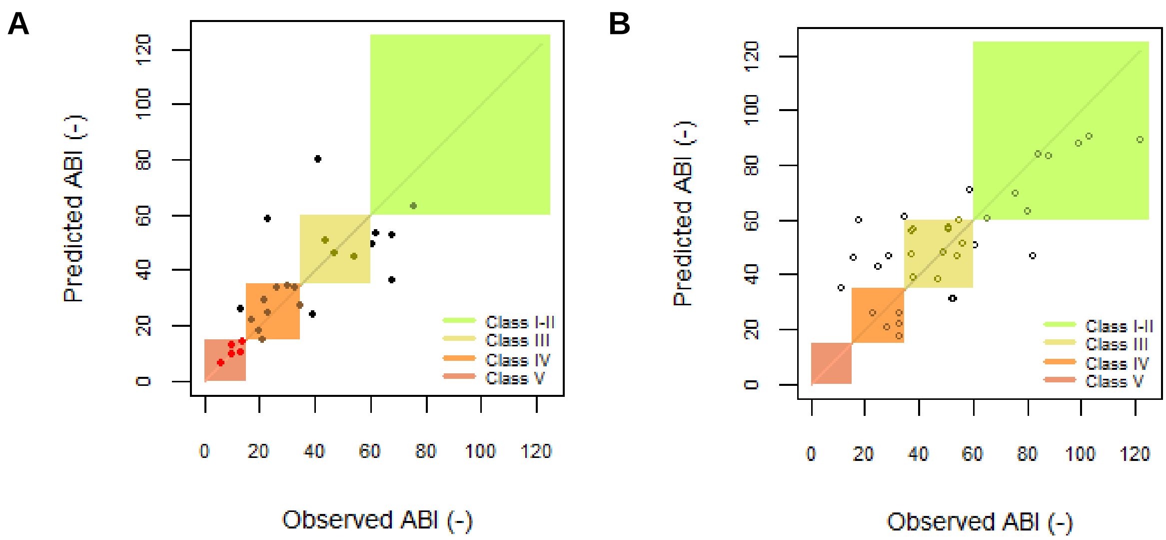


Fig. 6.3. Graph comparing observed data and simulated index obtained from the best models during the (A) dry season and (B) rainy season. The colors inside the squares indicate the biological water quality (ABI class). The dots inside the box are the sampling sites whose ABI classes were correctly predicted. 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.

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 6.4).

6.3.3 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, a comparison was done regarding 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 6.3) 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. E7.

Table 6.3. 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	

In addition, ammonium, DO and OS were selected in two of the four models developed using the complete dataset (Table E4), while, BOD_5, DO and OS were the explanatory variables in two of the models obtained for the without-outliers dataset (Table E5). In the main-rivers dataset, orthophosphate was a common predictor variable of the four models, while DO and OS were predictor variables for the ABI in three of the four constructed models (Table E6). Fig. 6.4A, Fig. E8A and Fig. E8B show the deviation of observed values and the predicted values with the complete, the without-outliers and the main-rivers datasets, respectively. Based on the three-fold cross-validation results obtained with the evaluation metrics: pseudo R^2, CCI and κ (Table 6.4), 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.

Table 6.4. 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

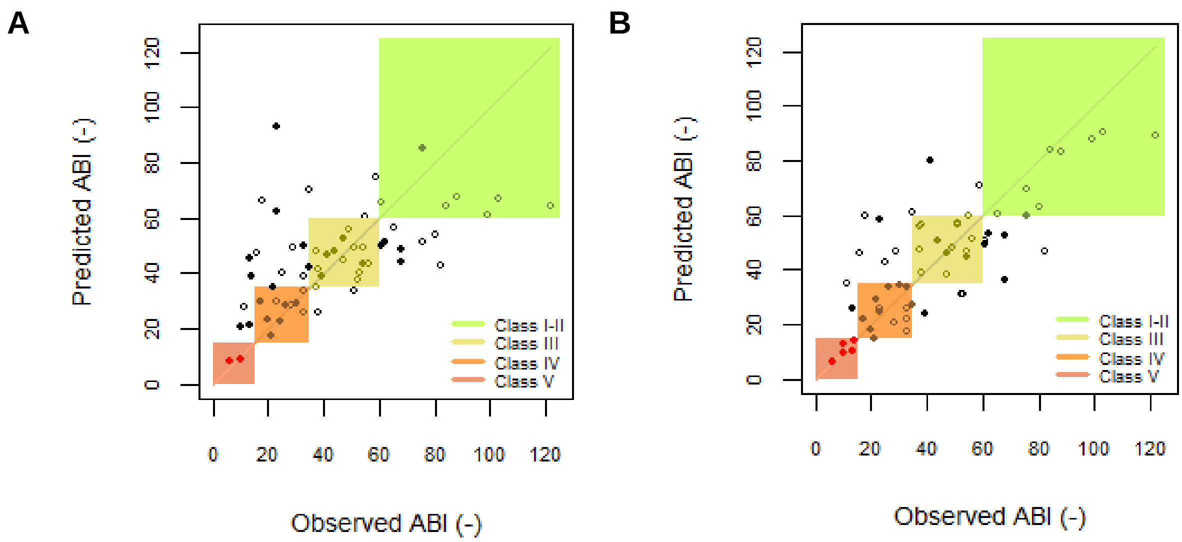


Fig. 6.4. Graph comparing simulated index and field data for models developed from: (A) the complete dataset, and (B) selective validation of the combination of the best models for the dry and the rainy seasons. The filled dots and the empty dots are the sampling sites taken during dry and rainy seasons respectively.

6

Based on the CCI and κ, the models trained with the different datasets were validated with the other datasets (Fig. 6. 2A). The best performing models (CCI~70% and κ >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 κ>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. 6.4B). 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 6.5 presents the validation results of models developed from the different datasets.

Table 6.5. 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		Model: Without-outliers dataset		Model: 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 ± 0.0
Without-outliers	54.9 ± 5.1	0.4 ± 0.1	51.6 ± 5.1	0.3 ± 0.1	52.1 ± 0.8	0.3 ± 0.0
Main-rivers	54.6 ± 4.9	0.3 ± 0.1	51.7 ± 6.0	0.3 ± 0.1	50.6 ± 1.1	0.3 ± 0.0
Dry-season	50.0 ± 4.4	0.3 ± 0.1	48.1 ± 5.0	0.3 ± 0.1	50.0 ± 3.1	0.3 ± 0.0
Rainy-season	58.6 ± 6.8	0.4 ± 0.1	54.3 ± 7.7	0.3 ± 0.1	53.6 ± 1.4	0.3 ± 0.0
Principal-river (Tomebamba River)	58.3 ± 2.4	0.4 ± 0.0	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 ± 0.0
Reference sites	37.5 ± 8.3	0.1 ± 0.1	39.6 ± 8.0	0.1 ± 0.1	35.4 ± 4.2	0.0 ± 0.1
Tributaries	52.8 ± 7.2	0.4 ± 0.1	52.8 ± 3.2	0.4 ± 0.0	52.8 ± 3.2	0.4 ± 0.0

Validation sets	Model: Dry-season dataset		Model: Rainy-season dataset		Selective validation of the 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	0.2 ± 0.0	47.6 ± 4.7	0.3 ± 0.1	59.3 ± 3.6	0.4 ± 0.1
Without-outliers	46.3 ± 2.1	0.2 ± 0.0	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	58.0 ± 2.9	0.4 ± 0.0
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 ± 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 ± 0.0	**60.7 ± 4.6**	**0.4 ± 0.1**	**69.1 ± 6.2**	**0.6 ± 0.1**
Urban sites	50.0 ± 1.9	0.3 ± 0.0	50.0 ± 12.5	0.3 ± 0.1	**65.3 ± 5.0**	**0.5 ± 0.1**
Reference sites	20.8 ± 4.8	-0.1 ± 0.1	33.3 ± 17.4	0.0 ± 0.2	33.3 ± 6.8	0.0 ± 0.1
Tributaries	43.1 ± 2.8	0.2 ± 0.0	45.8 ± 7.9	0.3 ± 0.1	**62.5 ± 9.5**	**0.5 ± 0.1**

6.4 Discussion

6.4.1 Biological and environmental differences between seasons in tropical rivers

One of the strengths of the study design of this chapter is the comparison between the combinations of season-specific models with season-overarching 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, it was discovered that there was a variation of macroinvertebrate assemblage between the dry and rainy seasons at the high altitude tropical environment (Jacobsen 1998).

6

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 by irrigation of agriculture and grassland areas in the suburban areas of the city (Jerves-Cobo et al. 2018b). 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 organic pollutants such as orthophosphate, BOD_5 and ammonium (Jacobsen 1998, Beyene et al. 2009). However, the depletion of oxygen concentration is influenced by these organic 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, Hoang et al. 2010b).

During the rainy season, the ABI on average was 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 TSol 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. 2019). 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 $mg.L^{-1}$ and 7.9 $mg.L^{-1}$. The average value of DO, obtained during the sampling 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 80-90%, 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 (OS) 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 the 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. 2018b). The total solids (TSol), 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 TSol 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 TSol 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 (TSol) 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).

6.4.2 Season-specific and season-overarching models

Our results demonstrate that GLMs can be used to select physicochemical variables that best predict the biological water quality based on the ABI index. However, in most models, κ 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.

6

The selective combination of season-specific models that were applied according to the season in which the sites were sampled, showed an accuracy of at least 5% and 10% higher in CCI and κ, respectively, than season-overarching models. 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. The higher concentration of organic pollutants and nutrients could have positively influenced the prediction of the ABI during dry 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. (2013a) 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 because shading and macrophytes were limited in the sampled sites (Jerves-Cobo et al. 2019). The validation results obtained from the rainy dataset were quite stable as indicated by the standard deviation of *κ,* with a variation between 0 and 10% of the mean *κ* (Table 6.5). 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 (Table E2 and Table E3).

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. 6

6.4.3 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 were not altered.

During the construction of the models, the results obtained from the datasets both with and without outliers were compared. An outlier is a result of either an influence observation or a measurement error (Rousseeuw and Leroy 2005). In this research, the influence observations were classified as outliers that were registered in the polluted tributaries of the main rivers. The models constructed in this chapter 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 (Table E4 and Table E5). 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. 2013b, Damanik-Ambarita et al. 2016). Moreover, to avoid possible overfitting in the linear models that could be produced with small datasets was followed 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).

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, contributed to the reliability and accuracy of the models showing relevant ecological patterns.

6.4.4 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. 2018b). For these reasons, the accuracy of the developed models to assess their applicability in improving water quality status and their effectiveness in river management was analyzed. 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, Machangaram 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.

6

6.5 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 modelling approach developed in this chapter can be applied and/or extrapolated to similar basins in tropical countries.

Chapter 7. Integrated ecological modelling for evidence-based determination of water management interventions in urbanized river basins: Case study in the Cuenca River basin (Ecuador)

Adapted from:

Rubén Jerves-Cobo, Lorenzo Benedetti, Youri Amerllnck, Koen Lock, Chaim De Mulder, Jana Van Butsel, Félipe Cisneros, Peter Goethals and Ingmar Nopens (submitted). Integrated ecological modelling for evidence-based determination of water management interventions in urbanized river basins: Case study in the Cuenca River basin (Ecuador).

Abstract

The growth of urbanization worldwide has contributed to the deterioration of the ecological status of water bodies. Efforts at improving the ecological status have been made either in isolated form or integrated measures by stakeholders, but in many cases, these measures have not been evaluated to determine their benefit. In this chapter, a scenario analysis is implemented to restore the ecological water quality in the Cuenca River and its tributaries, which are located in the southern Andes of Ecuador. For this analysis, an integrated urban wastewater system (IUWS) model, which gave satisfactory results in its calibration and validation processes, was linked with ecological models. The IUWS is a mechanistic model that incorporated the river water quality model, a wastewater treatment plant (WWTP) with activated sludge technology, and discharges from the sewage system. The ecological status of the waterways was evaluated with the Andean Biotic Index (ABI), which was predicted using generalized linear models (GLMs). The GLMs were calculated with physicochemical results from the IUWS model. Four scenarios that would enhance the current ecological water quality were analyzed. In these scenarios, the inclusion of a new WWTP with carbon, as well as with carbon and nitrogen removal, and the addition of retention tanks before the discharges of combined sewer overflows (CSOs) were assessed. The new WWTP with carbon and nitrogen removal would bring about a better restoration of the ecological water quality due to better nitrogen removal. The retention tanks would help to enhance the ecological status of the rivers during rainy seasons. The integrated model implemented in this chapter was shown to be an essential tool to support decisions in the Cuenca River basin management.

7.1 Introduction

The growth of urbanized areas around the world has increased pressure on their nearby aquatic ecosystems, which include rivers, estuaries, and lakes. Generally, these pressures originate in the management of the wastewater and runoff during rainfall. Thus, point discharges from outfalls and wastewater treatment plants (WWTPs) or combined sewer overflows (CSOs), and surface water outfalls (SWOs) during rain events disturb the ecological equilibrium of the receiving water body. This disturbance affects both the chemical composition of the water and the biotic structure (Hvitved-Jacobsen 1982, Mulliss et al. 1997) and contributes to the deterioration of the morphology of the waterways such as river beds, flow changes, and bank structures (Walsh et al. 2005).

To date, stakeholders have implemented isolated or integrated measures to enhance the water quality of the streams. These actions have been addressed to comply with national or local regulations that in most cases control the physicochemical composition of the water (Holguin-Gonzalez et al. 2013b). However, the optimization of resources invested in improving water quality requires planning tools. In the case of municipal water management, a good planning tool for this optimization may be the implementation of an integrated urban wastewater system (IUWS) model that includes WWTPs, sewage networks and receiving water. The IUWS model combines cost-efficient solutions, the analysis of possible synergies and the optimization of the wastewater system performance (Benedetti et al. 2013). The final outputs of an IUWS model are the concentration of organic pollutants and some hydrological variables in rivers such as water flows. This integrated model has demonstrated great potential in the scenario analysis to improve the river water quality (Deksissa et al. 2004, Solvi et al. 2005, Benedetti et al. 2007, Holguin-Gonzalez et al. 2014).

The ecological quality of freshwater bodies is affected by both physicochemical changes and hydro-morphological variation produced by natural and anthropogenic stressors (Burneo and Gunkel 2003, De Pauw et al. 2006). To preserve the aquatic ecosystems, national policies have introduced the concept of an ecological status in river management (Griffiths 2002, USEPA - U.S. Environmental Protection Agency 2011). This status includes an integrated assessment of the biological, hydro-morphological and physicochemical elements of the water quality (Griffiths 2002). In this context, ecological models have been developed as tools that can support environmental management and policy development. These types of models have provided an understanding of the influence of different pressures such as the impact of wastewater and CSOs, in addition to supporting the wastewater treatment selection or predicting the ecological water quality (Goethals and Forio 2018). As such, the integration of hydro-morphological aspects with physicochemical water quality has been applied as a tool in basin management to predict the possible restoration of ecological water quality (Tomsic et al. 2007, Mouton et al. 2009b). In urban areas, similar approaches have been developed to predict the final ecological status of the rivers (Holguin-Gonzalez et al. 2013b, Holguin-Gonzalez et al. 2014).

Since 1984, the Water Supply and Wastewater Management Municipal Company ETAPA-EP has been working on improving the water quality status of the Cuenca River and its tributaries. This effort has included biological water quality assessment as well as the use of a mathematical modelling approach to predict final physicochemical water quality; results that have supported the water management implementation (ETAPA-EP 2007). Holguin-Gonzalez et al. (2013a) evaluated the contribution of the Ucubamba wastewater treatment plant (U-WWTP) to the ecological water quality of the Cuenca River during the dry season. This assessment was executed by means of the linkage of the physicochemical water quality results provided by the QUAL2K model with ecological models. Although the water quality has been improved, the biological water quality varies between deficient and bad in most of the stretches of the urban area of the Cuenca River and its tributaries during the dry season; a quality that changes from moderate to good during the rainy season (Jerves-Cobo et al. 2019). The leading causes that affect the ecological status of the Cuenca River system are management problems of the wastewater and runoff such as leakage from the sewage system, suburban population that discharge directly into streams, the capacity of the U-WWTP, CSOs discharges, and disperse pollution from livestock areas (Greeley & Hansen and ACSAM 2017, Jerves-Cobo et al. 2018b).

The framework and objectives of this chapter are presented in Fig. 7.1. The aim of this chapter is to develop and to validate an integrated ecological model (IEM) for the analysis of possible scenarios to be applied to river restoration and management. This IEM assesses the variation in the ecological water quality produced simultaneously by physicochemical pollution and hydro-morphological pressures. In this regard, the IEM incorporated four models: (1) a mechanistic model used to predict the physicochemical water quality in rivers; (2) an activated-sludge wastewater treatment plant model; (3) a sewer model that generates combined sewer outfalls; and (4) an ecological model to assess the ecological river water quality. This framework was applied to analyze potential measures in the restoration of the Cuenca River that is located in the southern Andes of Ecuador.

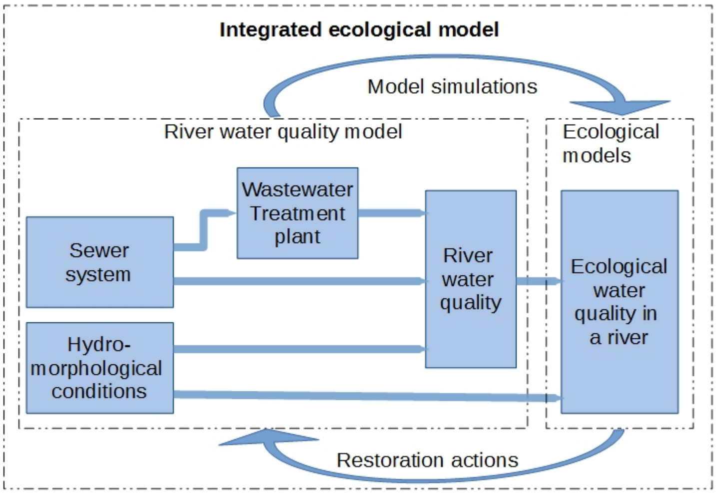


Fig. 7.1. Framework of the proposed integrated ecological modelling for scenario analysis in river restoration.

7.2 Integrated river water quality model: building, validation and implementation

In the simulation of the integrated river water quality system, an open model structure simulator called WEST® (Vanhooren et al. 2003) was used. WEST® has mainly been applied to the modelling and simulation of wastewater treatment plants, although this software also works with river water quality models and sewer models that are based on a cascade of continuously stirred tank reactors in series (CSTRS) (Connolly et al. 2004, Deksissa et al. 2004). The structure of WEST® is flexible enough to allow for the inclusion of a relevant process for the study or exclusion of a less critical process. Thus, in the first instance, the river water quality model was developed for the Tomebamba and Cuenca Rivers, so when this model was calibrated during the dry and rainy seasons (cf. section 7.2.1), it was included the elements shown in Fig. 7.2 for the scenario analysis. In the analysis, it simulated a new wastewater treatment plant during the dry season (Fig. 7.2A) as well as the new WWTP and tanks to control overflows from the combined sewer network from the city of Cuenca during rainfall events (Fig. 7.2B). Finally, to understand the ecological water quality in the stretches of the Tomebamba and Cuenca Rivers, the information obtained from the WEST® simulations was entered into the equations that link the physicochemical and morphological variables with the ecological water quality that was achieved in the ecological model (cf. Section 7.2.4).

7.2.1 River water quality model

For modelling the river, the River Water Quality Model No.1 (RWQM1) (Reichert et al. 2001) was used that can be integrated as a sub-model in the WEST® software. The river model was constructed with 20 CSTRs in series that simulates 27.1 km of the Tomebamba and Cuenca Rivers from Cuenca and its surrounding areas. Correspondingly, their urban tributaries such as the Tarqui (0.3 Km) and Machangara (0.5 Km) Rivers. These were also simulated with one CSTR each, while the Yanuncay

River (3.2 Km) was simulated with two CSTRs. Moreover, the small streams, the Amarillo River, El Tejar, Saucay, El Valle, and the Milchichig and Sinincay Brooks were included in the IEM as point discharges (Fig. 7.2A).

A

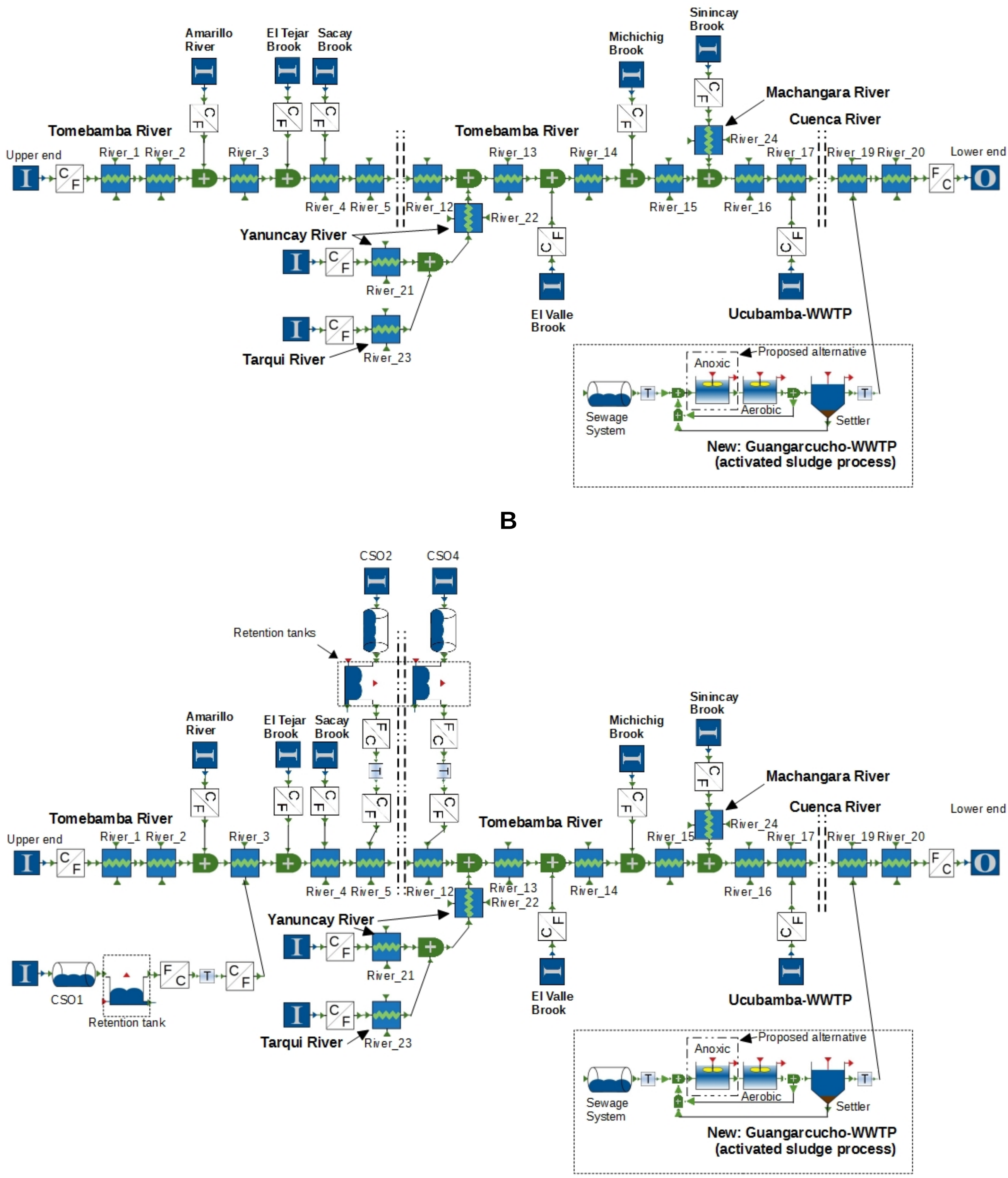


Fig. 7.2. Scheme of the water quality model of the actual conditions and proposed scenarios (inside dotted rectangles) in WEST®: (**A**) during dry season and (**B**) during rainy season.

7

The concentration of pollutants measured in the two sampling campaigns of 2015 and 2016 (cf. Section 2.2.1) was mostly computed as model state variables. The physicochemical information of the Cuenca River described in Section 2.2.1 was complemented with nine samples taken before and after the discharge of the U-WWTP in 2017 (Espinoza Berrezueta and Zumba López 2018). The variables in this analysis included dissolved oxygen (S_{O2}), ammonium-nitrogen (S_{NH}), nitrite-nitrogen (S_{NO2}), nitrate-nitrogen (S_{NO3}) and phosphate (S_{PO}). However, the fractions of model components concerning the COD were derived from the measurements, according to equations 7.1 and 7.2 (Reichert et al. 2001). These derived compounds were readily biodegradable soluble COD (S_S), inert soluble COD (S_I), particulate inert COD (X_I), particulate organic matter (X_S), heterotrophic biomass (X_H), first stage nitrifying bacteria (X_{N1}), second stage nitrifying bacteria (X_{N2}) and algae and macrophytes (X_{ALG}). The values of the fractional compounds of COD are shown in Table F1 and were taken from previous model applications (Benedetti 2006, Solvi 2006).

$$COD_{Total} = COD_{soluble} + COD_{particulate} \tag{7.1}$$

$$COD_{Total} = (S_S + S_I) + (X_I + X_S + X_H + X_{N1} + X_{N2} + X_{ALG}) \tag{7.2}$$

7

The soluble COD represented an average of 78% of the total COD; a value that was the result of the relationship between volatile dissolved solids and total volatile solids obtained from all measurements sampled during the dry and rainy seasons. Chlorophyll-a was transformed into algae and macrophytes using a factor recommended by Jorgensen et al. (1991). The relationship between the COD and the five-day biological oxygen demand (BOD_5) in the rivers showed a wide range of variation, changing from around three in polluted sites to almost 20 in reference sites. For that reason, the variation of this relationship (COD/BOD_5) was calibrated with a logarithmic regression (Fig. F1) that had an R^2 of 0.62. This relationship was applied to calculate the BOD_5 in each stretch of the Tomebamba and Cuenca Rivers.

The Tomebamba and Cuenca Rivers model was calibrated in two steps: the first step was the process of hydraulic calibration and the second step was the physicochemical calibration. The hydraulic calibration was performed by modifying the Manning roughness coefficient for the riverbed. This coefficient was chosen according to the range of values recommended by Te Chow (1959). Thus, the Manning roughness coefficient was set with an initial value that was subsequently adjusted to obtain the best fit of the model. For that, the values of flow and water depth predicted by the model were compared with the measured values. For the hydraulic calibration, the flow information provided by ETAPA-EP was used, which included measurements from 2015 to 2017 at six gauging stations located in the study area (Fig. 2.1). The flow in each gauging station was calculated with its respective equations which link water flow and water depth (ETAPA-EP 2015). In other river stretches, it was used the flow and the cross-river sections obtained in the two sampling campaigns of 2015 and 2016 as noted in section 2.2.1. Additionally, the flow information and the cross-river section were completed with measurements achieved in four sites on nine occasions during 2013. The slope and the cross section from site To12 to To19 in the Tomebamba River (Fig. 2.1) were obtained from the existing topography (PROMAS-UCuenca 2005). For the other stretches of the Tomebamba, Yanuncay, Tarqui and Machangara Rivers, the slope was calculated using the height of each site.

To calibrate and to assess the uncertainty of the river water quality model, the Monte Carlo analysis was used, in which each set of parameters that were part of the RWQM1 and that could influence the different variables (DO, COD, BOD_5, NH_4, NO_2, NO_3 and PO_4) was evaluated. The results of this analysis also generated confidence bands for the model, using the concepts of the Generalized Likelihood Uncertainty Estimation (GLUE) methodology (Beven and Binley 1992). For this calibration, a local sensitivity analysis was first conducted (Saltelli et al. 2004). This analysis was performed in three different stretches located upstream from the Tomebamba River in the urban area. For each stretch, the five parameters selected that influenced each of the variables were those whose mean central relative sensitivity (MeanCRS) had the highest absolute values. If the three analyzed stretches had different parameters per variable, in total a maximum of 15 parameters per variable was selected from the local sensitivity analysis. Next, those chosen parameters per variable underwent a global sensitivity analysis (Meirlaen 2002, Vanrolleghem and Gillot 2002), which was also applied to the three upstream stretches. From the latter analysis, the five parameters whose standardized regression coefficient (SRC) had the highest absolute value were selected for the scenario analysis. Thereafter, in the scenario analysis, these five parameters were evaluated from approximately 1,000 possible combinations uniformly distributed in specific ranges, choosing the values of the parameters that gave the best goodness-of-fit for each variable (Table F2). For this selection, in the first place, a visual inspection was applied to evaluate the quality of the model. Finally, the minimum root mean square error (RMSE) of each variable and the set of variables in the analysis was calculated in addition to the square of the Pearson’s Product Moment Correlation Coefficient (R^2) and the values of chi-squared (χ^2) (Mac Berthouex and Brown 2002). The following physicochemical variables were calibrated: dissolved oxygen, COD, BOD_5, ammonium, nitrite, nitrate, phosphate (PO_4), average water depth, flow velocity and roughness. The validation was accomplished in three stretches located downstream in the Cuenca River. The calibration and validation were done during both the dry and rainy seasons.

7.2.2 Wastewater treatment plants

The wastewater treatment plant of the Ucubamba (U-WWTP - Fig. 2.1), which has been in operation since 1999, processes an average flow of 1.8 $m^3.s^{-1}$. The U-WWTP is divided into two identical flow lines, each comprising a line of an aerobic pond, a facultative lagoon and a maturation pond (Alvarado et al. 2012). The discharge from the U-WWTP was also included in the integrated model (Fig. 7.2) to understand its impact on the river water quality, which involved the average flow and concentration of pollutants during the dry and rainy seasons of 2015 and 2016; values that were calculated from the weekly records provided by ETAPA-EP.

A new wastewater treatment plant will be built in the Guangarcucho (G-WWTP) region, located downstream from Cuenca. This new G-WWTP will increase the current capacity of the wastewater treatment, and will include the flow from the suburban areas that will be connected into the urban sewage system. The G-WWTP was designed for carbon removal only with activated sludge technology and it was included in the integrated model in WEST® (Fig. 7.2). The steady-state simulation of the G-WWTP was performed on the basis of the flow (1.2 $m^3.s^{-1}$) and pollutant concentrations, as well as kinetic parameters provided by ETAPA-EP, which were used in its design (Greeley & Hansen and ACSAM 2017). For the dynamic simulation of the G-WWTP, the hourly and daily variation of the flow and concentration of pollutants were calculated on the basis of the information obtained in the existing U-WWTP (Durazno 2013). As a model scenario, the G-WWTP was upgraded (Fig. 7.2) in order to include ammonium removal. The processes of the G-WWTP were simulated with the Activated Sludge Models No. 1 (ASM1) that includes carbon oxidation in aerobic and anoxic conditions as well as nitrification and denitrification (Henze et al. 2000). The upgraded G-WWTP was analyzed in both steady and dynamic states with the same flow and charges used for its design.

7.2.3 Discharges from the combined sewer overflow

In the integrated model, four discharges from the combined sewer overflows (CSOs) were included, which represents 40 CSOs existing along the two banks of the Tomebamba River.

The variation of the pollution concentration and flow in the discharges of the CSOs were calculated with information collected in three overflows located along the Tomebamba River, during nine rainfall events in the rainy seasons of 2017 and 2018 (Schrijver 2017, Montalvo et al. 2018). Similarly, the flow of the CSOs according to their contribution area was calculated, the surface runoff coefficient obtained for the city of Cuenca by Rubio et al. (2017) and precipitation from rainfall events. Three simulations were obtained from rainfall events with their variation on flow and pollutants in the discharges of the CSOs.

In the scenario analysis, four retention tanks before the discharges of the CSOs were included into the Tomebamba River. These tanks reduce both the amount of water spilled from the CSOs, and the peak of pollutant concentrations that affect the water quality in the rivers. Furthermore, the retention tanks allow for the storage of the pollutant water from the CSOs during rainfall and conduction to the WWTP after the storms (Marsalek et al. 2014). The volume of the retention tanks was calculated with a runoff from a 10-year (EPA 1999), one-hour storm, obtained with the rational method for areas smaller than 162 ha (NYC Environmental Protection 2012). For this volume, a detention time of 12 minutes was also considered, which provided a storage of 20 m^3/ha, a value that was in the range of typical storage in Germany (Zabel et al. 2001).

7.2.4 *Ecological Model - GLM model: relation between chemical water quality and invertebrate index (ABI)*

Ecological models were developed to determine association parameters, which could influence the performance of the Andean Biotic Index (ABI) in the study area. The methodology to obtain the ecological model as well as the final models were described in Sections 2.3.3 and 6.2.

From the models developed in Chapter 6, two models were selected for application in the dry season: one model wasused in streams with a low concentration of pollutants, while a second model was used in stretches with a higher pollution concentration. In the rainy season, a different model was chosen for the prediction of the ABI Class in the Tomebamba and Cuenca Rivers.

7.2.5 *Scenario analysis for restoration of the ecological water quality*

In order to understand how the new G-WWTP and other actions could improve the ecological water quality in different stretches of the Tomebamba and Cuenca Rivers, four different scenarios were simulated (Fig. 7.2) with the calibrated RWQM during the dry and rainy seasons, which are summarized in Table 7.1. The implementation of the G-WWTP implies that the capacity to handle wastewater treatment will be increased. Consequently, the pollution from suburban areas that is currently discharging into small streams, brooks and main tributaries could then be connected into the sewage system and conducted to the wastewater treatment system. The total population that could be connected to the G-WWTP was taken from Greeley & Hansen and ACSAM (2017). However, this population was distributed along suburban and rural areas where discharges are occurring directly into streams.

Table 7.1. Scenarios to recover the ecological water quality in the Tomebamba and Cuenca Rivers.

Scenario	Season	Actions
Sc-1	Dry season	Implementation of the new G-WWTP (carbon removal).
Sc-2	Dry season	Implementation of the upgraded G-WWTP (carbon and nutrients removal).
Sc-3	Rainy season	Implementation of the New G-WWTP (carbon removal).
Sc-4	Rainy season	Implementation of the upgraded G-WWTP (carbon and nutrients removal).
Sc-1 to Sc-4	Dry and rainy seasons	Additional actions to be included in Sc-1 to Sc-4: Reduction in the concentration of pollutants in 80% of small streams and brooks: Amarillo River and El Tejar, Saucay, El Valle, Milchichig and Sinincay Brooks, due to the connection of the direct discharges with the urban sewage system. Reduction in the concentration of pollutants in 50% of the main effluents: Yanuncay, Tarqui and Machangara Rivers, due to the connection of the direct discharges with the urban sewage system.
Sc-3 & Sc-4	Rainy season	Additional actions to be included in Sc-3 and Sc-4: Implementation of four retention tanks before CSO discharges.

7.3 Ecuadorian water quality standards

The Ecuadorian government has established regulations regarding the preservation of the aquatic ecosystems in freshwater (MAE-Ecuador 2015). The standard norm set a value limit for different parameters to control organic pollution. Thus, the thresholds of DO, nitrate, nitrite, phosphorous, BOD_5, turbidity and total ammonium nitrogen are shown in Table 7.2. In order to comply with the aforementioned regulation, these threshold limits must not be exceeded.

Table 7.2. Ecuadorian regulations to control dissolved oxygen, organic pollutants and nutrients in freshwater aquatic ecosystems.

Parameter	Units	Value	Comment
Dissolved Oxygen	% of saturation	>80%	
Nitrate (NO_3)	$mg \cdot L^{-1}$	13	
Nitrite (NO_2)	$mg \cdot L^{-1}$	0.2	
nitrogen- total phosphorous		A ratio of 15: 1	The value was taken for waters of aesthetic use
Five-day biological oxygen demand (BOD_5)	$mg \cdot L^{-1}$	2 – 6	Aquatic ecosystems with a moderate impact
Total ammonium (NH_4)	$mg \cdot L^{-1}$		Concentrations are regulated concerning the pH and temperature

Temperature (°C)	pH 6.5	7.0	7.5	8.0	8.5	9.0
10	32.4	10.30	3.26	1.04	0.34	0.12
15	22.0	6.98	2.22	0.72	0.24	0.09
20	15.2	4.82	1.54	0.50	0.17	0.07

7.4 Results

7.4.1 River water quality model

The river water quality model was calibrated and validated during both seasons for hydraulic and chemical variables, displaying reliable predictions. The hydraulic calibration and validation showed a high determination coefficient (R^2>0.7) for water depth and flow velocity (Table F3). The physicochemical variables presented a different goodness-of-fit in the calibration and validation processes during both seasons. Fig. 7.3 and Fig. F2 present the graphs of the calibration and validation for the different variables (DO, BOD_5, COD, ammonium, nitrite, nitrate and orthophosphates) along the Tomebamba and Cuenca Rivers during the dry and rainy season, respectively, along with their confidence bands of the 5th and 95th predicted percentiles, obtained from the Monte Carlo analysis. Similarly, Table F3 presents the values of R^2, RMSE and χ^2 obtained with physicochemical variables during the dry and rainy seasons in the calibration and validation processes.

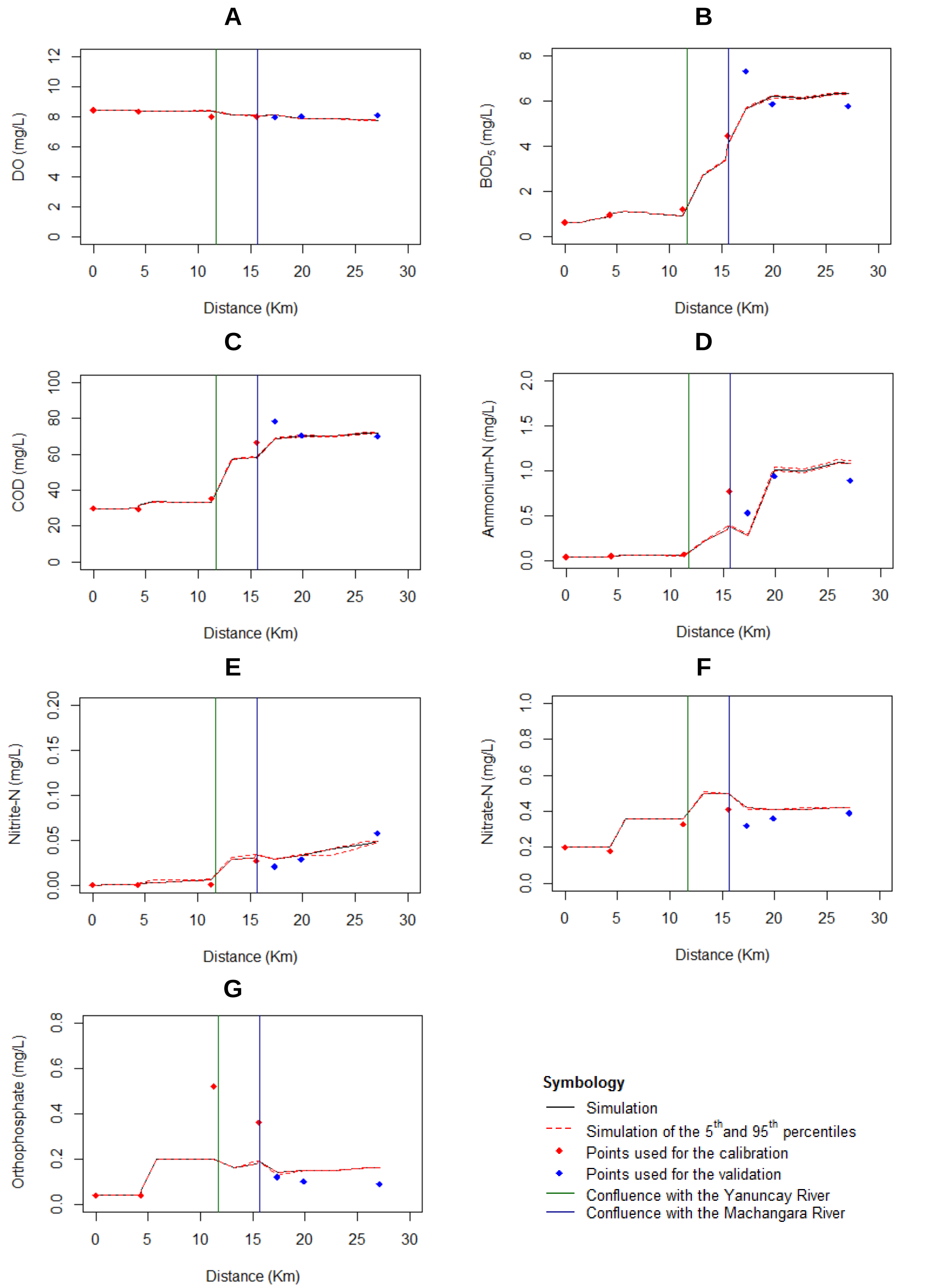


Fig. 7.3. Calibrated water quality model in the Tomebamba and Cuenca Rivers during the dry season for: (**A**) DO, (**B**) BOD_5, (**C**) COD, (**D**) Ammonium, (**E**) Nitrite, (**F**) Nitrate, (**G**) Orthophosphate.

For the calibration during the dry season, the BOD_5, COD, ammonium, nitrite, nitrate and orthophosphate had values higher than 0.7, while the DO showed a lower value (R^2<0.45). For the validation, nitrate was the only variable with low values (R^2<0.45), the other variables had values higher than 0.7. Similarly, for the calibration during the rainy season, BOD_5, ammonium and nitrate had an R^2 higher than 0.7, while nitrite presented a moderate determination coefficient and the DO, COD and orthophosphates had low values (R^2<0.45). The validation in this season displayed different results than those that were obtained during the dry season. Thus, BOD_5, nitrate and nitrite had values higher than 0.7, while ammonium and orthophosphate indicated moderate values (0.45≤R^2≤0.7) and the other variables displayed low determination coefficients.

7.4.2 Ecological assessment model

The more reliable ecological model to predict the ABI class was obtained when one model per season was selected. Namely, the best performing GLM models during both the dry and rainy seasons were chosen according to their highest κ, CCI and R^2 obtained during the validation process. As a result, the ABI class had different predictors per season (Table 7.3). Thus, for the dry season, the ABI had three predictors: BOD_5, NH_4 and PO_4, while for the rainy season the six variables that influenced the ABI were: NO_2, NO_3, DO, oxygen saturation, total solids and bank material. The GLMs chosen belong to the Gamma and Gaussian families. When these two chosen GLMs were combined and selectively applied according to their season to the Tomebamba and Cuenca Rivers, the CCI and κ had values higher than 70% and 0.4, respectively (Table 7.3).

Table 7.3. The best-performing season-specific models and their selective combination (Selection from Chapter 6 - Jerves-Cobo et al. (2019))

Explanatory Variables	Regression Parameters	Season-specific models			Selective combination of season-specific models	
		Dry-season Gamma models:		Rainy-season Gaussian model:	Combination of the rainy season + dry season models:	
		mD1.3fcv2a3	mD8_2.3fcv1a2	mR8.gaussian	mR8.gaussian + mD1.3fcv1a3	mR8.gaussian + mD8_2.3fcv1a2
		Coefficient [a]	Coefficient [a]	Coefficient [a]		
	A	1.3E-02	1.3E-02	-725.5		
Nitrate	B1		3.5E-02	85.3		
Nitrite	B2			-1100.7		
Ammonium	B3	-4.2E-02	6.2E-03			
BOD_5	B4	5.2E-03	1.2 E-03			
DO	B5			-42.8		
Oxygen Saturation	B6			10.9		
Orthophosphate	B8	1.4E-01				
Total solids	B10			-0.1		
Bank material	B18			10.6		
Training subset (2/3)						
AIC:		140.6	133.86	314.0		
Pseudo R^2:		0.7	0.7	0.6		
CCI:		72%	47%	69%		
κ:		0.6	0.3	0.5		
Validation subset (1/3)						
Pseudo R^2:		0.4	0.1	0.3		
CCI:		63%	22%	73%		
κ:		0.4	0.2	0.6		
Applied to the Tomebamba and Cuenca Rivers datasets						
CCI		57%	29%	86%	76%	67%
κ:		0.4	0.1	0.7	0.7	0.5

[a] The coefficient is multiplied by its variable respectively within the GLM equation.

7.4.3 Integrated ecological model and scenario assessment

Actions to be considered in order to improve biological water quality involve reducing the concentration of organic pollutants in the receiving water body. Consequently, the fulfillment of the Ecuadorian regulation to preserve the aquatic ecosystem was evaluated after the implementation of different scenarios. Thus, in Sc-1 and Sc-2 constructed for dry season and in Sc-3 and Sc-4 developed for rainy season (Table 7.1), BOD_5, ammonium, nitrite and nitrate would remain under their thresholds (Table 7.2 and Fig. F3) in the Tomebamba and Cuenca Rivers. Furthermore, with the proposed measurements in Sc-1, Sc-2, Sc-3 and Sc-4, the Tomebamba and Cuenca Rivers will have an important decrease in the concentration of the analyzed pollutants: BOD_5, ammonium, nitrite, nitrate and orthophosphate, than in the current registered concentrations (Fig. F3). However, in Sc-1 and Sc-2 in the Tarqui River, its level of BOD_5 would remain over the regulated thresholds during the dry season, despite the fact that this river included a 50% reduction in pollutants.

The scenarios analyzed for the restoration of the ecological water quality have positive impacts in different stretches of the Tomebamba and Cuenca Rivers and their tributaries as can be seen in Fig. 7.4. Thus, with the implementation of the new G-WWTP, the capacity of the wastewater treatment will be increased, connecting the suburban areas of the city to the urban sewage system, and eliminating the direct sewage discharges into the waterways. With these measures in the dry season, the Tomebamba and Cuenca Rivers (Sc-1) and their streams would ensure the maintenance of the ABI class in the moderate category upstream from the city of Cuenca until the discharge of the current U-WWTP. From this point forward, the ABI class would remain in the deficient range. However, Sc-2 shows that by removing nutrients in the new G-WWTP, the ABI class would remain deficient, but with values closer to the moderate class. During the rainy season, in which the Tomebamba and Cuenca Rivers and their tributaries would have better water quality than during the dry season, the measures included in the Sc-3 such as retention tanks before the CSOs, would also show improvements in the biological water quality. The good ecological water quality of the Tomebamba River would continue until its confluence with the Machangara River. Additionally, the good water quality would be restored in the last three kilometers of the Cuenca River in the Sc-4. The upgrade to the new G-WWTP to include carbon and nutrients removal would positively influence the ABI class during both the dry and rainy seasons.

The retention tanks before combined sewage overflows (CSOs - Sc-3 and Sc-4) were tested during three rain events, from which one had a runoff from the 10-year, one-hour storm. These tanks showed effectiveness in diminishing pollutants discharged into the Tomebamba River by at least 50%, with the possibility of a reduction up to 100% as long as the polluted storage water is conducted to the WWTPs following precipitation.

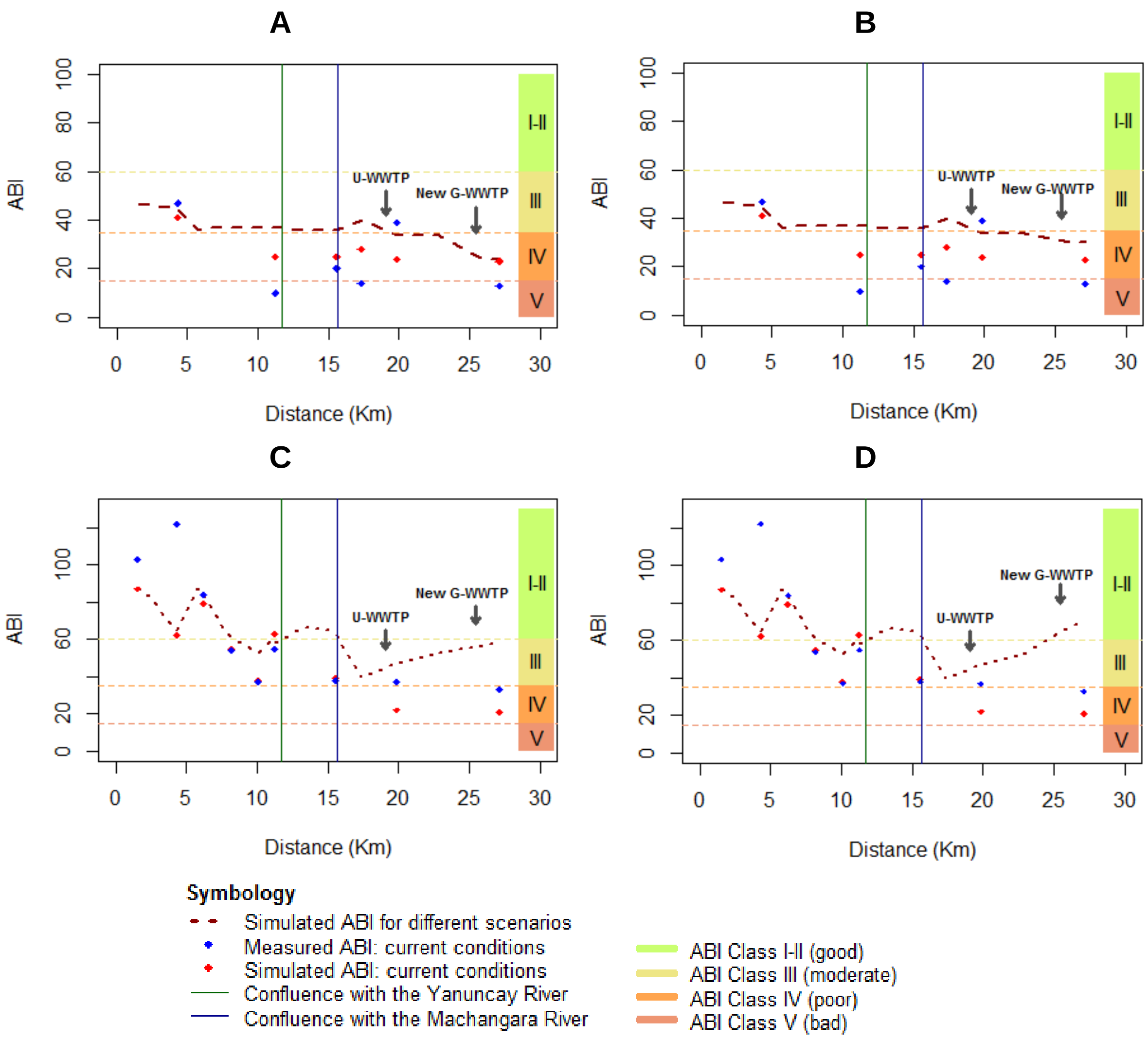


Fig. 7.4. Scenario analysis concerning restoration needs and expected outcomes: (A) scenario 1 (Sc-1) – dry season, (B) scenario 2 (Sc-2) – dry season, (C) Sc-3 (rainy season) and (D) Sc-4 (rainy season). The difference between Sc-1 and Sc-2 and between Sc-3 and Sc-4, is the technology used in the WWTP-G, thus Sc-1 and SC-3 are only with carbon removal, while for Sc-2 and Sc-4 they are with carbon and nutrients removal.

The analysis of possible technologies to be applied to the new G-WWTP demonstrated that the carbon and nutrient removal technology (Sc-2 and Sc-4) would contribute more effectively to the restoration of the Cuenca River's water quality than with the carbon removal technology (Sc-1 and Sc-3). Thus, for example, when compared with the new G-WWTP with carbon and nutrients removal with only carbon elimination, BOD_5, COD and total suspended solids would have the same levels of effectiveness. However, under steady-state conditions, the upgraded technology (carbon and nutrients removal) would be more efficient in approximately 70% of the removal of ammonium and approximately 60% in the elimination of nitrate, total nitrogen and total Kjeldahl nitrogen (TKN). In order to obtain the nitrogen species removal, the new G-WWTP must include an anoxic zone in its biological reactors

7

reducing its flow capacity. An analysis of the solids retention time (SRT) and hydraulic retention time (HRT) in comparison with the nitrogen and carbon families' removal was applied to the new G-WWTP in Fig. 7.5 and Fig. F4, respectively. In these graphs, it can be noted that with an SRT of 11 days and an HRT of 0.45 days, increased removal of ammonium and TKN was obtained, while higher values of the SRT or HRT did not influence the elimination of BOD_5 and COD.

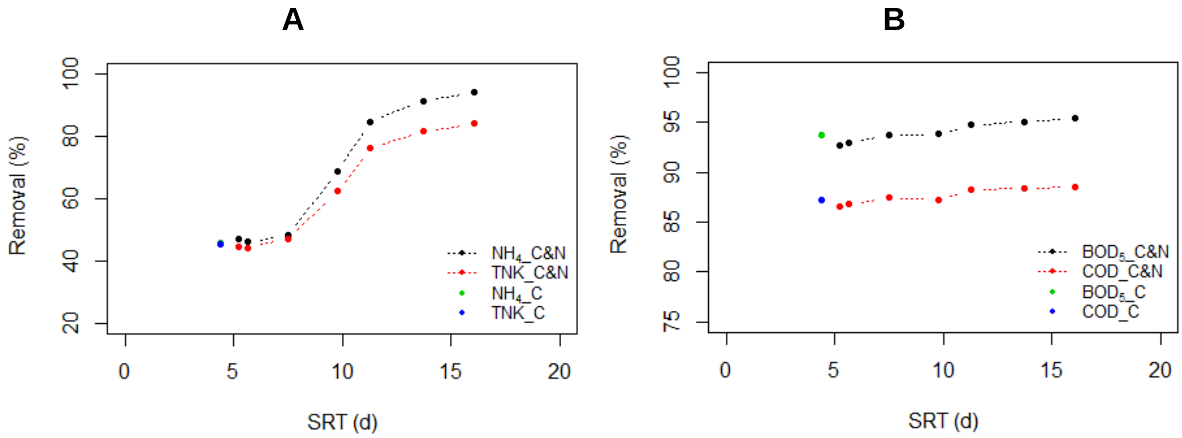


Fig. 7.5. Analysis of removal for the new G-WWTP according of solids retention time (SRT) with the technologies of carbon (C) and carbon and nitrogen elimination (C&N) for: (**A**) nitrogen family, and (**B**) carbon family.

7

7.5 Discussion

7.5.1 River water quality model, performance, uncertainty and validation

The primary goal for the calibration of the river water quality model was to have a base with enough accuracy that reflects the behavior of the rivers and a model that can be used as a planning tool to simulate scenario analysis to improve current water quality conditions. The model was calibrated under steady-state analysis during the dry and rainy seasons. This kind of calibration was accomplished with the data, which was taken in both sampling campaigns of 2015 and 2016, but did not consider the variation of the water quality at the sampled sites during a specific day or week. During the calibration process, the parameters set for RWQM1 were adjusted to minimize the difference between measurements and model predictions. Thus, the variables such as COD, BOD_5, NH_4, NO_3 and PO_4 presented acceptable goodness-of-fit with the default values of parameters set on the RWQM1 (Reichert et al. 2001). Only significant parameters were adjusted for this calibration: the re-aeration coefficient (Kla_{base}). This parameter is part of the re-aeration rate and was used to calibrate DO concentration during both seasons resulting in different values for each season. The re-aeration rate is also related to stream velocity, temperature and water depth; variables that registered a relevant change between seasons and likely influenced the different value of the Kla_{base} obtained in the calibration of the DO. In a preliminary calibration, the maximum specific growth rate of 2nd stage nitrifiers ($K_{gro,N2}$), which means the growth of organisms that oxidize nitrite to nitrate, was also included for the adjustment of NO_3

during the dry season. $K_{gro,N2}$ that influences the growth of 2nd–stage of nitrifiers could modify the values of the DO and phosphates as well as the particulate nitrogen (X_N_2) (Reichert et al. 2001). However, with the new value of $K_{gro,N2}$ the goodness-of-fit of the DO diminished, which is why this parameter was discarded. The BOD_5 and COD presented acceptable accuracy in both calibrations developed for the dry and rainy seasons. However, the constant relation between these variables (BOD_5/COD = 0.35), included in the RWQM1, was changed for each river stretch according to the registered values. In a study carried out in northwestern Nigeria, Mustapha et al. (2012) found a non-linear relationship between BOD_5 and COD during the dry season. In the northern Andes of Ecuador, Burneo and Gunkel (2003) identified that the ratio between COD/BOD_5 varied from 1.1 to 7.3 during the rainy season and from 1.5 to 2.9 during the dry season; values that were in the range obtained in this research. This ratio variation range between the COD/BOD_5 in the rivers could have been due to the difference in carbon sources, such as municipal versus industrial pollution in urban areas, or in the rural areas, organic matter from the moorland areas with a decreased biodegradability. Regarding the biodegradability of dissolved organic matter in moorland areas, Stutter et al. (2013) found that this ranged from 5 to 19% of the total transported dissolved organic matter.

The range of errors probably for both the calibration and validation depend on the variables. Thus, according to the R^2 values obtained in the calibration processes, the DO was low for the dry and rainy seasons, but its RMSE value was very low during both seasons. Similarly, the R^2 of COD during the rainy season was in the low range in the calibration and validation processes, but its RMSE was lower than other variables that presented a higher R^2. Orthophosphates also demonstrated a low value of R^2 and the highest RMSE in its calibration process given during the rainy season. The difference of R^2 between variables and seasons could also be influenced by the limited data available. In this regard, Achleitner (2008) pointed out that the R^2 could be insensitive to consistent and an over or underestimation of the predicted values. Furthermore, according to the GLUE technique, the results of the model calibration indicated a good prediction capacity, showing that variables were mainly in the range of the 95% confidence bands. Although the R^2 could likely be increased, changing the values of other parameters could make the model over-parameterized due to the limited dataset and given adequate goodness-of-fit with a different set of calibration parameters (Brun et al. 2001). For future research in which a dynamic calibration may be developed, it is recommended that various sampling sites should be sampled simultaneously with continuous hourly data along a defined period of time. This new information could help to improve the values of the calibration obtained in this research.

The new G-WWTP, which will be constructed in the near future, was not calibrated, because it was included in the integrated model with the same parameters used in its design. This new G-WWTP was designed with the ASM1 using the software GPS-X (Greeley & Hansen and ACSAM 2017). It is recommended to calibrate the ASM1 used in the integrated model with information that will be collected from the new G-WWTP, when this enters into production. This calibration would allow for the analysis of possible scenarios to improve the performance of the G-WWTP and supply information to the integrated model constructed in this chapter. Similarly, the operation of retention tanks was analyzed with data from three of nine rainfall events collected, from which one had a return period of 10-year. However, for future analysis, in which a different return period could be included, it is suggested to calibrate the flow caused by the rainfall events in conjunction with their registered pollution.

7.5.2 Ecological model, performance, uncertainty and validation

To predict the biological water quality (ABI) obtained under the four scenarios proposed, the information resulting from the RWQM1 given by WEST®, was included in the ecological models. For that reason, season-specific models were selected that had good fitting, measured according to their κ, CCI and R^2 (Jerves-Cobo et al. 2019). The implementation of two models in dry season along the Tomebamba and Cuenca Rivers revealed better accuracy than with the application of only one model. The first model had better precision in its application with a low concentration of pollutants — the Tomebamba River until its confluence with the Yanuncay —, while the second model performed better in streams with higher pollution — in the rest of the Tomebamba and the Cuenca Rivers. The accuracy of the ecological models (first model of the dry season and the rainy season model) was acceptable, with a κ higher or equal to 0.4, a CCI that fluctuates between 63% and 73%, and an R^2 in the range of 0.4 to 0.7. However, the second model applied in the dry season had a lower accuracy in total, yet it was more precise in the prediction of sites with higher pollution. When the chosen season-specific models were combined and selectively applied according to their season to the Tomebamba and Cuenca Rivers, the accuracy of the combined model was good, with a κ and CCI higher than 0.4 and 70%, respectively. With regard to ecological models, Mac Nally (2000) pointed out that the average R^2 of these models is 0.5. The size of the dataset could have influenced the accuracy of the ecological model, so 27 and 35 samples were applied to the construction of these models in dry and rainy seasons, respectively. However, according to Stockwell and Peterson (2002), a dataset whose size is higher than 50 samples has an accuracy close to the maximum, and if this size is diminished to 10 observations, the accuracy is 90%. In addition, ecological models represent a simplified view of reality, which implies that models always contain errors in assumptions, formulation and parameterization (Lek 2007). The season-specific models were more accurate in the prediction of the ABI class than the season-overarching models. This is likely due to the fact that the main variables to predict the ABI were different between seasons (Jerves-Cobo et al. 2019). These variables changed according to their concentration measured per season; thus, a weaker correlation between biotic indices and organic pollutants was obtained during the rainy season (Jacobsen 1998). The models applied in this chapter have demonstrated that they have relevant ecological patterns. Thus, usage of GLMs and the standardized sampling helped to obtain reliable and accurate models. However, the results obtained with ecological models applied in this chapter will give a good prediction as long as the range of the variables is the same as those used in the construction of the models (McCullagh 1984).

The dissolved oxygen (DO), organic pollutant expressed as BOD_5 and nutrients are widely acknowledged to have influence on the presence/absence of different families of macroinvertebrates (Thorne and Williams 1997, Jacobsen et al. 2003, De Pauw et al. 2006). Total solids (TSol) and bank material were two variables that were part of the ecological model, but that were not predicted by the RWQM1. The concentration of the TSol, which was used to forecast the ABI, was calculated with a simple mass balance, in which the possible sedimentation in the bed of the rivers, the increment by erosion and the loss for reactions were omitted. For the bank material, the classification registered in the survey that was applied to collect the information (Jerves-Cobo et al. 2018b), included bedrock, boulder, cobble, pebble, gravel, sand, silt and clay. Regarding the TSol, Gray and Ward (1982) found that this variable influenced the community structure of macroinvertebrates. Similarly, bank material, a morphological characteristic, was also found to affect invertebrate communities (Burneo and Gunkel 2003).

7

7.5.3 Integrated ecological modelling and scenario analysis

Two kinds of techniques can be handled in the integration of the models for river management: the integration and the combination approaches (Lam et al. 2004). The integration approach was used to join the river model (RWQM1) with the WWTPs (ASM1) and with the sewage systems (Reichert et al. 2001). This integration allowed for the prediction of the water quality compounds, the flow velocity and the depth into receiving rivers. As such, this approach has been addressed with acceptable accuracy in several studies (Meirlaen et al. 2001, Deksissa et al. 2004, Benedetti et al. 2007). Similarly, the combination approach was applied to link the results from the RWQM1 with ecological models obtained in the dry and rainy seasons. This combination was implemented to simulate both the current conditions and to analyze scenarios to restore the environmental state of the Cuenca River and its tributaries. Comparable applications to analyze restoration scenarios, in which a river water quality model was linked with ecological models, have been developed with good performance by Mouton et al. (2009b) and Holguin-Gonzalez et al. (2014).

The implementation of different scenarios for the restoration of the water quality in the Cuenca River system demonstrated their effectiveness in the analysis presented in this research. Thus, the implementation of the new G-WWTP will bring improvements to the biological water quality in Cuenca's River system, collecting wastewater from households located in the suburban area of the city of Cuenca. However, with the measures analyzed during the dry season, it was possible to obtain a moderate water quality in the Tomebamba River until its confluence with the Yanuncay River and in the Cuenca River from the U-WWTP onwards. The water quality of the Yanuncay River is affected by the Tarqui River, which also has an impact on the Tomebamba River. In the same way, the Machangara River influences the water quality conditions of the Cuenca River. In this way, the quality of a water body is affected by the quality of its tributaries (Ward and Stanford 1995). The primary cause of degradation of the Tarqui River is due to the diffuse fluxes of organic pollution from an extensive livestock area (Beltrán et al. 2013, Jerves-Cobo et al. 2018b). For the restoration of the Tarqui River, a future study can analyze measures to recover the biological water quality in livestock areas. Measures such as the implementation of buffer strips in each bank of the rivers could be applied (Mouton et al. 2009b). Regarding the Machangara River, point discharges from industries must be controlled to enhance the water quality in this tributary and subsequently in the Cuenca River.

The technologies of carbon removal or carbon and nutrients removal that could be implemented in the new G-WWTP will comply with the regulation for discharges to water bodies for nutrients, BOD_5, and COD (MAE-Ecuador 2015). However, the water quality in the Cuenca River during the dry season would not improve to moderate, due to the effluent from the current U-WWTP, which has a concentration of ammonium that varied between 10 to 20 $mg.L^{-1}$ and its BOD_5 is around 30 $mg.L^{-1}$ (ETAPA-EP 2017). The implementation of a new G-WWTP with carbon and nitrogen removal technology (Sc-2) rather than with only carbon removal technology (Sc-1) would elevate the biological water quality, but its class would remain deficient in the last analyzed stretch of the Cuenca River during the dry season. This improvement in the biological water quality obtained with a WWTP with carbon and nitrogen removal technology (Sc-2) is due to the higher efficiency in the ammonium and TKN removal, which allows a lower concentration of these pollutants to enter the Cuenca River. Regarding the activated sludge with extended aeration plants, a technology used in the design of the new WWTP (Greeley & Hansen and ACSAM 2017), McCarty and Brodersen (1962) indicated that the major problem with this technology is the lack of an appropriate nitrification process, which causes poor ammonium removal. Moreover, if the new G-WWTP could operate in anoxic and aerated conditions to remove ammonium, its limitation would be the solids retention time (SRT) as well as the hydraulic retention time (HRT). The SRT and HRT factors must be increased to guarantee the nitrogen removal, which would diminish the flow to be processed, an aspect that could likely be managed only within the first few years of plant operation. The new G-WWTP could also be enhanced with the implementation of an anaerobic zone for improving biological phosphorous removal. This analysis of the G-WWTP could be done under the Activated Sludge Model No. 2d (ASM2d) (Henze et al. 2000).

Although orthophosphates, a part of total phosphorous, is a component of the ecological model, this pollutant was not included in this study. This was because, with the removal of orthophosphate achieved by the new G-WWTP under Sc-1 and Sc-2, the concentration of orthophosphates in the Cuenca River was increased in 0.01 $mg.L^{-1}$ by their discharge (Fig. F3). Furthermore, if the removal of orthophosphate in the G-WWTP could be increased to 90%, the effluent of this WWTP would not change the concentration of this pollutant in the Cuenca River. This slight rise in orthophosphates would not affect the predicted ABI class. Consequently, the GLM used to calculate the ABI class in dry season was not sensitive to a slight variation of orthophosphates. In addition, if the removal of orthophosphate in the G-WWTP could be increased to 90%, the effluent of this WWTP would not change the concentration of this pollutant in the Cuenca River, which means the ABI class would be the same.

The scenarios analyzed to improve the biological water quality during the rainy season (Sc-3 and Sc-4) showed that the upgrade of the new G-WWTP from carbon removal to carbon and nitrogen removal, would have an influence in the improvement of the ecological water quality in the last studied stretch of the Cuenca River. In fact, in Sc-4, the Andean Biotic Index (ABI) would rise to good, although its value would not significantly change in comparison with the ABI obtained in Sc-3. In this regard, Jacobsen (1998), in a study developed in the northern Andes of Ecuador, identified a weaker relationship between chemical parameters and biotic indices during the rainy season. This was likely due to lower concentrations of pollutants as a consequence of both a higher dilution addressed by a higher flow in the river that reduced the effects of the discharge from the WWTP as well as the higher velocity of the water that generated a washing out of the organic material accumulated in the substrate (Jacobsen 1998, Burneo and Gunkel 2003, Jerves-Cobo et al. 2018b). Yet, as previously mentioned, during the dry and rainy seasons, the new G-WWTP with either carbon removal or carbon and nitrogen removal technology, the water quality in the Cuenca River system would improve. At the same time, the rivers would comply with the Ecuadorian regulations to preserve the aquatic ecosystem.

For rainy season, an analysis of the inclusion of the retention tanks (Sc-3 and Sc-4) before the discharges of the combined sewer overflows (CSOs) was included. Thus, from the CSOs, at the beginning of a rainfall event, high peak concentrations of dissolved compounds are generally discharged into rivers. The concentration of these discharges varies from an equal level found in the sewage system; this level increases in approximately the same factor as the CSO flow rises (Holzer and Krebs 1998). As a result, the aquatic ecosystem of the receiving water is affected by a short-term and delayed impact in the DO depletion, non-ionized ammonium and shear stress (Hvitved-Jacobsen 1982, Borchardt and Sperling 1997). Furthermore, discharges from CSOs could cause extended periods of anoxic sediment in the river's bed (Hvitved-Jacobsen 1982, Reichert et al. 2001). The retention tanks as shown in section 7.4.3

would undoubtedly increase the biological water quality and diminish the impact of the first flush, as well as, the amount of dissolved pollutants discharged directly into the rivers. In addition, these structures would help to reduce the quantity of sediment that is discharged to the rivers. In this regard, Erbe et al. (2002) pointed out that combined water retention tanks perform positively in relatively short events with medium intensity, but for high intensity events that produce the strongest wash-out effects, these tanks do not have a significant influence. It should be noted that with the volume of the tanks that were the lowest recommended range of storage sized (Zabel et al. 2001), it was obtained a good performance with a rainfall event of a return period of 10-year and a duration of one hour, reducing the amount of pollution that could impact the river by 50%. Consequently, to improve the water quality in the urban area of the Cuenca River system, the inclusion of retention tanks before the discharges of CSOs is recommended. It is also suggested to test the proposed retention tanks with stronger rainfall events than were analyzed in this chapter.

The results obtained in this chapter will enable stakeholders such as ETAPA-EP and the National Secretary of Water (SENAGUA), to gain insight on how the proposed measures are interrelated with the ecological status of the Cuenca River and its tributaries. Further analysis to enhance the water quality especially in the branches of the Cuenca River, in which the DPSIR (drivers, pressures, state, impact and response) framework would be adopted, can be developed in the near future.

7.6 Conclusion

The four scenarios applied to restore the ecological water quality in the Cuenca River system, which was measured with the Andean Biotic Index (ABI), demonstrated that the proposed measures in this chapter would help to enhance the quality of this resource in the urbanized area of the Cuenca River and its tributaries. Thus, with the implementation of the new G-WWTP, the benefits in the improvement of the ABI would be most significant during the dry season, with either carbon removal or carbon and nitrogen removal technologies. However, the carbon and nitrogen removal technology would bring a higher restoration to the ecological status of the Cuenca River due to the increased nitrogen removal during both seasons. The retention tanks before the discharges of combined sewer overflows (CSOs) would also improve the ecological water quality in the urban area of the Cuenca River system during rainy seasons. The scenarios analyzed in this chapter would enable stakeholders to gain insight prior to implementing measures to improve the water quality conditions of the Cuenca River system. In order to accomplish this, the implementation of the new G-WWTP is a priority for this purpose. Finally, similar applications with the construction of an integrated ecological model could be replicated in other basins to analyze the impact of different measures on the ecological water quality in other water bodies.

Chapter 8: Conclusion and perspective

8.1 Introduction

The results obtained in each of the chapters of this thesis give insights into the current conditions of the water quality of the Cuenca River system, and the potential measures that can be taken for its restoration.

Here, the major take home messages in view of the posed research questions are presented as well as some general thoughts on the future direction of the integrated river management model that can be implemented in future studies.

8.2 Research questions analysis

8.2.1 Biological impact assessment of sewage outfalls in the urbanized area of the Cuenca River basin

- What is the current biological water quality in the urban and suburban areas of the Cuenca River?

Two biological indexes were used to calculate the biological water quality and to assess the impact of sewage outfalls in the urbanized area of the Cuenca River basin: the Andean Biotic Index (ABI) and the Biological Monitoring Working Party adapted to Colombia (BMWP-Col) were applied. The biological water quality varied between bad to very good for both indexes, showing better water quality during the rainy season. This better water quality was due to a higher presence of the sensitive taxa Ephemeroptera—Plecoptera—Trichoptera (EPT). The water quality degraded when the river passed through the city, exhibiting a better water quality in sites located upstream.

- What is the impact of the sewage outfalls in the urbanized area of the Cuenca River basin?

Generally, the water quality was better upstream from outfalls, where there is a higher presence of sensitive taxa than downstream. This situation was similar upstream and downstream from a discharge of combined sewer outfalls (CSO) during the rainy season.

- Which factors influence the water quality in the urban and suburban areas of the Cuenca River basin?

The variation of the biological indexes was mainly related to oxygen saturation, organic pollutants and nutrient concentrations. Thus, organic pollutants and nutrients in the rivers had higher concentrations during the dry season - with the exception of nitrite - than during the rainy season. During the rainy season, the water velocity and turbidity in the rivers were higher, while the dissolved oxygen (DO) was lower. However, the concentrations of the DO during both seasons was close to the saturation level, which in the high Andes region is relatively low in comparison with low altitude sites.

- Is the biological water quality influenced by different variables in each season?

The influence of organic pollutants and nutrients were evident during the dry season. While, during the rainy season this relationship was weaker, and a higher influence on dissolved oxygen and morphological characteristics of the rivers could be detected.

- Which biological index is more suitable for the high Andean region?

Both the ABI and the BMWP-Col showed similar patterns. However, the results suggest that the ABI is more suitable than the BMWP-Col for the high Andean region to assess water quality (above 2400 m a.s.l.), because of the more relevant score outcomes.

8.2.2 A Methodology to Model Environmental Preferences of Ephemeroptera-Plecoptera-Trichoptera Taxa in the Machangara River basin (Ecuador)

- What is the water quality in the Machangara River basin, a sub-basin of the Cuenca River?

The Machangara River is an affluent of the Cuenca River, in which its flow is regulated throughout the year by the presence of two dams. The biological water quality, expressed as the BMWP-Col, varied along the Machangara River revealing poor water quality both after the dams and in the city of Cuenca. This poor water quality was related to increased levels of organic pollutants and extremely high flow velocities during hydropower discharges. 8

- Which physicochemical and morphological variables are related to the occurrence of sensitive taxa - the Ephemeroptera—Plecoptera—Trichoptera (EPT) taxa?

Logistic regression models, which are a family of generalized linear models (GLMs), were applied to understand the habitat preferences and to predict the presence/absence of the three families of the EPT taxa: *Baetidae*, *Leptoceridae and Perlidae.* The explanatory variables of the constructed GLMs differed substantially among the taxa. In total, eight variables: chemical oxygen demand (COD), five day-biological oxygen demand (BOD_5), conductivity, temperature, true color, flow velocity, fecal coliforms (FCs), and pH had a substantial influence in the presence of the aforementioned sensitive taxa.

- How were the sensitive taxa affected by different physicochemical variables, affected including those with different values of these variables?

Each taxon was affected by different physicochemical variables. Thus, *Baetidae* was influenced by the COD, conductivity, temperature and true color. Whereas, *Leptoceridae* was related to the BOD_5, COD, conductivity, flow velocity, FCs and pH. Finally, conductivity, flow velocity, FCs, and temperature influenced the presence of *Perlidae*. Although some explanatory variables were similar between families, their preference ranges were different.

8.2.3 Model-Based Analysis of the Potential of Macroinvertebrates as Indicators for Microbial Pathogens in Rivers

- What is the current microbial water quality in the Machangara River basin?

The Machangara River showed that the microbial water quality had a variation from upstream to downstream, in which the influence especially of cattle areas and of the suburban and urban areas was significant. The latter areas have higher concentrations of fecal coliforms (FCs) than upstream sites.

- Is the biological index, BMWP-Col, influenced by microbial pathogens?

It was found that the biological index BMWP-Col was not related to the fecal coliforms, mostly because extreme conditions were not encountered during the sampling campaigns.

- Can the presence or abundance of sensitive taxa of macroinvertebrates be influenced by microbiological water quality?

To obtain this possible relationship between sensitive taxa of macroinvertebrates with microbiological water quality, habitat-suitability models based on the decision trees models (DTMs) were implemented. The DTMs linked the *Baetidae*, *Scirtidae* and *Perlidae* sensitive taxa, with the thresholds of fecal coliforms that are given by two Ecuadorian water use regulations. Thus, some sensitive taxa such as *Baetidae*, *Scirtidae* and *Perlidae* could be influenced by microbial pathogens.

- Can sensitive macroinvertebrates be used as proxy indicators in the fulfillment of the microbiological water quality regulations?

Two DTMs were obtained that could be used as proxy indicators in the fulfillment of the microbiological water quality regulations. Thus, an abundance higher than four for *Baetidae* and five for *Scirtidae* could be an indicator of poor water quality that cannot be used for either irrigation or for activities with primary contact. People with minimal training to identify these families of macroinvertebrates could use this technique as a proxy and a preliminary indicator of fecal pollution in river water.

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

- How well do ecological models perform in one season versus another season?

Different models were developed to link the Andean Biotic Index (ABI) with physicochemical and morphological variables. These river ecological assessment models (EAMs) were constructed with specific information about a season, and with all the data that was collected during the dry and rainy seasons. For the construction of these models, three kinds of generalized linear models (GLMs): Gaussian, Gamma, and Inverse Gaussian were applied. The information collected in the urbanized area of the Cuenca River basin was the basis for the construction of the models. The only variable removed for the model's construction was temperature, which was correlated with dissolved oxygen. The results revealed low accuracy when the models constructed in one season were tested in a different season. Even, the ABI had different predictor variables per season. However, the season-specific models constructed for the dry season were more reliable than those developed for the rainy season. Consequently, patterns between ecological water quality and organic pollutants had better performance during the dry season.

- What is the benefit of season-overarching models?

The season-overarching models can indicate the systematic influence of variables in both seasons. However, they cannot reflect the difference found between seasons. 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. The added values depends on the use of the models and available data.

- Is the use of a selective combination of season-specific models more effective than the use of a season-overarching model?

The selective combination of season-specific models was more accurate than season-overarching models. As such, season-specific models provide the key variables that influence the biological water quality in each particular season. The sites where the data was collected influenced the models' construction. Thus, both the season-specific models and season-overarching models were more accurate in urban areas, where the majority of the samples were collected than in pristine sites. Hence, these results provide the pattern to key actions for each season that could be implemented by stakeholders to restore the water quality of the Cuenca River. An ecological model is an important tool that can be applied in river restoration. However, ecological models can 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.

8.2.5 Integrated ecological modelling for evidence-based determination of water management interventions in urbanized river basins: Case study in the Cuenca River basin

- Can various measurements be simulated in an integrated urban wastewater system?

The Cuenca River was simulated in an integrated urban wastewater system (IUWS) model, which included a river model (RWQM1), the discharges from the sewage system, and a wastewater treatment plant (WWTP), as well as, a new WWTP with activated sludge technology (ASM1). The IUWS model was calibrated and validated in a static condition, for the following variables: COD, BOD_5, NH_4, NO_3, and PO_4. It was performed with information collected during the dry and rainy seasons. Furthermore, to understand the impact of each proposed measure in the ecological water quality of the Cuenca River, the results obtained with the IUWS were linked with the ecological models developed in Chapter 6.

- Which measures are more effective to improve the current water quality in the urban and suburban areas of the Cuenca River basin?

In the IUWS model, four scenarios were analyzed to improve the water quality in the urban and suburban areas of the Tomebamba and Cuenca Rivers. From which, the connection of isolated sewage systems to the urban sewage network, as well as, the construction of a new wastewater treatment plant (WWTP) with activated sludge technology will play an important role in the restoration of the water quality during both seasons. Knowledge of the current water quality of the Cuenca River and the variables that influence its ecological status, in conjunction with the results of the proposed scenarios to restore the aquatic ecosystem, would enable stakeholders to gain insight into how implement measures that enhance the ecological water quality of the Cuenca River and its affluents.

- Does the use of retention tanks located before the discharges of the combined sewer overflows (CSOs) influence the water quality of the rivers?

The inclusion of retention tanks before the discharges of the CSOs will help to improve the water quality of the Cuenca River during rainy seasons. These tanks will diminish the impact of the first flush and the amount of dissolved pollutants discharged from the CSOs into the rivers during rainy events. Furthermore, the retention tanks allow storage of the polluted water from the CSOs during rainfall and conduction to the WWTP after the storms.

- What is the benefit of a wastewater treatment plant (WWTP) with carbon removal compared with a WWTP with carbon and nitrogen removal in the water quality of the Cuenca River?

A new WWTP with carbon and nitrogen removal will be more effective in the restoration of the water quality in the Cuenca River than the WWTP with only carbon removal. This is because, with less nitrogen discharging from the new WWTP, the Cuenca River will receive less pollution loads allowing for the improvement to its ecological water quality.

8.3 Recommendation for future research

This thesis has provided valuable insights that contribute to the river restoration of not only the Cuenca River basin, but also to watersheds worldwide. However, additional studies in different seasons and regions, and at alternate altitudes may contribute to and may provide new insights and information for a deeper understanding of ecological river restoration. This newfound knowledge could help to improve the development of an integrated river management (IRM) model. Below, proposed future studies are presented.

8.3.1 Sampling campaigns and dynamic data

Sampling campaigns to assess river water quality rarely measure the presence of substances such as metals, pesticides, and drugs, although these substances can contribute to the deterioration of the aquatic environment of water bodies. In order to understand the real impact of these substances on river ecology, it is recommended they be included in future sampling campaigns in the Cuenca River basin and basins worldwide.

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 produced by rainfall and irrigation. Consequently, the water quality of rivers displays a dynamic variation over time. However, this water quality is assessed with discrete sampling campaigns that offer insight mainly into the impact generated from natural and anthropogenic sources. For future studies in the Cuenca River basin, the installation of an automatic monitoring system is recommended, which could provide information about the physiochemical water quality variation over time. This new measurement equipment must be strategically located in the basin to collect water quality and quantity variation simultaneously. These measurements must be accompanied by at least one biotic monitoring per season. This biotic monitoring should continue with samples taken of macroinvertebrates that have demonstrated the capability to be adequate bio-indicators in the high Andean region. For future biotic monitoring, continued use of the Andean Biotic Index to assess the biological status of the rivers is recommended. Knowledge about the variation of pollutants concentration over time will allow for a better understanding of their impact on the ecological water quality variation. Furthermore, this dynamic and sequential information could help in understanding the implication of climate change on the ecological water quality of the rivers.

The inclusion of new measurements of substances such as metals, pesticides, and drugs as well as garnering continuous information on physicochemical water quality could provide new information for improving the river management of urban and rural environments. The latest information could then be utilized in the development of new ecological and integrated river management models. The results provided by these models may offer different perspectives for stakeholders in river management.

8.3.2 Area of the Cuenca River basin to be analyzed

The primary goal of this dissertation was to understand the current situation of the water quality in two areas of the Cuenca River basin with the highest anthropogenic impact. However, a third area of the Cuenca River basin, which corresponds to the Tarqui River sub-basin, in which land usage is primarily for livestock will also need to be studied. The influence of diffuse organic pollution on water quality and how its impact could be handled is a pending issue. However, this issue could help to improve the water quality not only in the Tarqui River, but in the Cuenca River as well. In cattle areas, the water quality was degraded by nutrient and sediments transported by runoff from rainy events and irrigation. For that reason, a future study in the Tarqui River basin, the inclusion of buffer zones along the riparian area (Anbumozhi et al. 2005) of the Tarqui and its affluents may be a solution to consider. Additionally, the effect of the width of the buffer zones in the water quality recovery could be included in new studies (Parkyn et al. 2005). The recovery of the ecological status of streams surrounded by small livestock areas located in the Cuenca River basin could also be part of any future analysis.

In addition, these possible measures to restore the water quality of the Tarqui River could be included in the Soil and Water Assessment Tool (SWAT) software. SWAT is a semi-distributed basin model, and flexible and open-source software, in which the cattle and agriculture areas, as well as households and habitat modifications can be included (Douglas-Mankin et al. 2010, Arnold et al. 2012). However, to understand the benefit of the proposed scenarios in the ecological water quality, the results generated by SWAT will need to be linked with ecological models.

8.3.3 Integrated urban water system components

The integrated water system (IUWS) model implemented in this dissertation provided valuable information for urban wastewater management and its impact on river water quality. However, the IUWS model could be improved with future research.

Thus, the new WWTP, which will be constructed with activated sludge technology, was analyzed with two sub-technologies: carbon removal, and carbon and nitrogen removal. However, in a future analysis of this new WWTP, a study of the phosphorus removal and its implication on both the WWTP and the recovery of the Cuenca River water quality could be included. Additionally, when the new WWTP is constructed, the IUWS model, developed in this dissertation, could be calibrated with the information that could be collected in the initial operation of this new facility. Moreover, a calibrated WWTP model could help to analyze different scenarios in the operation of this new facility that will provide information to operators on how to deal with routine and extreme conditions in the flow and quality of the wastewater.

The information from the CSOs used in this dissertation was provided by a parallel study (Montalvo et al. 2018). This study contributed information about the flow and pollutants variation during the operation of a CSO. The pollutants analyzed were COD, BOD_5, nitrite, nitrate, ammonium, and orthophosphates. However, not all of these pollutants were measured during registered rainfalls. As such, in a future study, it is recommended to take simultaneous measurements of flow and the aforementioned pollutants during CSO discharges in separate rainy events. Furthermore, turbidity and total solids measurements should also be included in this study. Moreover, the study could focus on understanding the flow and water quality variation produced by different rainy events and dissimilar areas of contribution.

Another recommendation for potential research is an analysis of the benefits of the application of urban best practice (BMPs) in reducing the impact of CSOs on the water quality of the rivers. These BMPs can be combined with low-impact development (LID) procedures (Dietz 2007, Damodaram et al. 2010), which could further reduce their flow and improve the discharged CSO water quality. The analysis of the BMPs and LID procedures can then be proposed, undertaken and analyzed using the Storm Water Management (SWM) model (Rossman and Huber 2016).

8.3.4 Model uncertainty

In this dissertation, the ecological models, the integrated urban wastewater (IUW) model, and their link into the integrated river management (IRM) model have demonstrated and provided essential and unique insights for the development of river management plans. Nevertheless, the integration of various models with their specific data could also contribute to an increased uncertainty of the IRM model. Consequently, the rise in uncertainty adds to the increase in false positives, which may provide erroneous outputs. The new uncertainty of the IRM model was not studied in this Ph.D. dissertation. For this reason, it is suggested that in prospective studies that include model integration, the propagation of a model's uncertainty also be assessed. Thus, techniques such as Bayesian hierarchical modelling (BHM) or an integrated Bayesian hierarchical framework - techniques that have been used to handle hydrological uncertainty (Liu and Gupta 2007) - may be applied in the uncertainty analysis of integrated models.

References

Achleitner, S. 2008. Modular conceptual modelling in urban drainage development and application of city drain. IUP-Innsbruck University Press.

Acosta, R., and H. Hampel. 2015. Evaluación del estado ecológico y biodiversidad de macroinvertebrados bentónicos en la cuenca alta del río Paute y Parque Nacional El Cajas. Universidad de Cuenca, Cuenca - Ecuador.

Acosta, R., M. Rieradevall, B. Ríos, and N. Prat. 2009. Propuesta de un protocolo de evaluación de la calidad ecológica de ríos andinos (CERA) y su aplicación a dos cuencas en Ecuador y Perú. Limnetica **28**:035-064.

Acreman, M., and A. Ferguson. 2010. Environmental flows and the European water framework directive. Freshwater Biology **55**:32-48.

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

Agresti, A., and M. Kateri. 2011. Categorical data analysis. Springer.

Alvarado, A., S. Vedantam, P. Goethals, and I. Nopens. 2012. A compartmental model to describe hydraulics in a full-scale waste stabilization pond. Water Research **46**:521-530.

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

Ambelu, A., K. Lock, and P. Goethals. 2010. Comparison of modelling techniques to predict macroinvertebrate community composition in rivers of Ethiopia. Ecological Informatics **5**:147-152.

Ambelu, A., S. Mekonen, M. Koch, T. Addis, P. Boets, G. Everaert, and P. Goethals. 2014. The application of predictive modelling for determining bio-environmental factors affecting the distribution of blackflies (Diptera: Simuliidae) in the Gilgel Gibe watershed in Southwest Ethiopia. PloS one **9**:e112221.

Anbumozhi, V., J. Radhakrishnan, and E. Yamaji. 2005. Impact of riparian buffer zones on water quality and associated management considerations. Ecological Engineering **24**:517-523.

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

Ansa, E., H. Lubberding, J. Ampofo, G. Amegbe, and H. Gijzen. 2012. Attachment of faecal coliform and macro-invertebrate activity in the removal of faecal coliform in domestic wastewater treatment pond systems. Ecological Engineering **42**:35-41.

Aquastat, F. 2016. Water withdrawal by sector, around 2010. Food and Agriculture Organization of the United Nations. Last modified November.

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

Arnold, J. G., D. N. Moriasi, P. W. Gassman, K. C. Abbaspour, M. J. White, R. Srinivasan, C. Santhi, R. Harmel, A. Van Griensven, and M. W. Van Liew. 2012. SWAT: Model use, calibration, and validation. Transactions of the ASABE **55**:1491-1508.

Arnone, R. D., and J. P. Walling. 2007. Waterborne pathogens in urban watersheds. Journal of Water and Health **5**:149-162.

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

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

Ballestero, M., V. Arroyo, and A. Mejía. 2015. Universalización de servicios de Agua Potable y saneamiento. Page 105 *in* CAF, editor. VII World Water Forum Regional, Korea.

Beadle, L. C. 1974. The inland waters of tropical Africa. An introduction to tropical limnology. Longman Group Ltd, Publishers, 74 Grosvenor Street, London W. 1.

Beauchard, O., J. Gagneur, and S. Brosse. 2003. Macroinvertebrate richness patterns in North African streams. Journal of Biogeography **30**:1821-1833.

Beltrán, A., P. Mendieta, and J. Vanegas. 2013. Calidad del agua y contaminantes en el río Tarqui. Revista Galileo - Universidad de Cuenca **23**:115-121.

Benedetti, L. 2006. Probabilistic design and upgrade of wastewater treatment plants in the EU Water Framework Directive context. Ghent University.

Benedetti, L., J. Langeveld, A. Comeau, L. Corominas, G. Daigger, C. Martin, P. S. Mikkelsen, L. Vezzaro, S. Weijers, and P. A. Vanrolleghem. 2013. Modelling and monitoring of integrated urban wastewater systems: review on status and perspectives. Water Science and Technology **68**:1203-1215.

Benedetti, L., J. Meirlaen, F. Sforzi, A. Facchi, C. Gandolfi, and P. Vanrolleghem. 2007. Dynamic integrated water quality modelling: A case study of the Lambro River, northern Italy. Water SA **33**:627-632.

Bennett, L. E., and M. Drikas. 1993. The evaluation of colour in natural waters. Water Research **27**:1209-1218.

Beven, K., and A. Binley. 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrological processes **6**:279-298.

Beyene, A., W. Legesse, L. Triest, and H. Kloos. 2009. Urban impact on ecological integrity of nearby rivers in developing countries: the Borkena River in highland Ethiopia. Environmental monitoring and assessment **153**:461.

Boix-Fayos, C., G. Barberá, F. López-Bermúdez, and V. Castillo. 2007. Effects of check dams, reforestation and land-use changes on river channel morphology: case study of the Rogativa catchment (Murcia, Spain). Geomorphology **91**:103-123.

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

Borchardt, D., and F. Sperling. 1997. Urban stormwater discharges: ecological effects on receiving waters and consequences for technical measures. Water Science and Technology **36**:173-178.

Brooker, M., and P. Johnson. 1984. The behaviour of phosphate, nitrate, chloride and hardness in twelve Welsh rivers. Water Research **18**:1155-1164.

Brown, R. M., N. I. McClelland, R. A. Deininger, and R. G. Tozer. 1970. A water quality index- Do we dare.

Brun, R., P. Reichert, and H. R. Künsch. 2001. Practical identifiability analysis of large environmental simulation models. Water Resources Research **37**:1015-1030.

Burneo, P. C., and G. Gunkel. 2003. Ecology of a high Andean stream, Rio Itambi, Otavalo, Ecuador. Limnologica-Ecology and Management of Inland Waters **33**:29-43.

Cairns, J., and J. R. Pratt. 1993. A history of biological monitoring using benthic macroinvertebrates. Freshwater biomonitoring and benthic macroinvertebrates **10**:27.

Cameron, A. C., and F. A. Windmeijer. 1997. An R-squared measure of goodness of fit for some common nonlinear regression models. Journal of Econometrics **77**:329-342.

CEPAL. 2012. Diagnóstico de las estadísticas del agua en Ecuador. CEPAL.

Chakona, A., C. Phiri, and J. A. Day. 2009. Potential for Trichoptera communities as biological indicators of morphological degradation in riverine systems. Hydrobiologia **621**:155-167.

Chatterjee, S., and A. S. Hadi. 1986. Influential observations, high leverage points, and outliers in linear regression. Statistical Science:379-393.

Cohen, J. 1960. A Coefficient of agreement for nominal scales. Educational and Psychological Measurement **20**:37-46.

Connolly, N., M. Crossland, and R. Pearson. 2004. Effect of low dissolved oxygen on survival, emergence, and drift of tropical stream macroinvertebrates. Journal of the North American Benthological Society **23**:251-270.

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.

D'Heygere, T., P. L. M. Goethals, and N. De Pauw. 2003. Use of genetic algorithms to select input variables in decision tree models for the prediction of benthic macroinvertebrates. Ecological Modelling **160**:291-300.

Dakou, E., T. D'Heygere, A. P. Dedecker, P. L. M. Goethals, M. Lazaridou-Dimitriadou, and N. De Pauw. 2007. Decision tree models for prediction of macroinvertebrate taxa in the river axios (Northern Greece). Aquatic Ecology **41**:399-411.

Dallas, H. F., and J. A. Day. 2007. Natural variation in macroinvertebrate assemblages and the development of a biological banding system for interpreting bioassessment data—a preliminary evaluation using data from upland sites in the south-western Cape, South Africa. Hydrobiologia **575**:231-244.

Damanik-Ambarita, M. N., G. Everaert, M. A. E. Forio, T. H. T. Nguyen, K. Lock, P. L. S. Musonge, N. Suhareva, L. Dominguez-Granda, E. Bennetsen, and P. Boets. 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.

Damodaram, C., M. H. Giacomoni, C. Prakash Khedun, H. Holmes, A. Ryan, W. Saour, and E. M. Zechman. 2010. Simulation of combined best management practices and low impact development for sustainable stormwater management 1. JAWRA Journal of the American Water Resources Association **46**:907-918.

De Pauw, N., B. Beyst, and S. Heylen. 2000. Development of a biological assessment method for river sediments in Flanders, Belgium. Internationale Vereinigung für theoretische und angewandte Limnologie: Verhandlungen **27**:2703-2708.

De Pauw, N., W. Gabriels, and P. L. Goethals. 2006. River monitoring and assessment methods based on macroinvertebrates. Pages 113-134 Biological monitoring of rivers. John Wiley and Son, Ltd., Chichester, UK.

Dedecker, A. P., P. L. Goethals, W. Gabriels, and N. De Pauw. 2004. Optimization of Artificial Neural Network (ANN) model design for prediction of macroinvertebrates in the Zwalm river basin (Flanders, Belgium). Ecological Modelling **174**:161-173.

Degraer, S., E. Verfaillie, W. Willems, E. Adriaens, M. Vincx, and V. Van Lancker. 2008. Habitat suitability modelling as a mapping tool for macrobenthic communities: an example from the Belgian part of the North Sea. Continental Shelf Research **28**:369-379.

Deksissa, T., J. Meirlaen, P. J. Ashton, and P. A. Vanrolleghem. 2004. Simplifying dynamic river water quality modelling: A case study of inorganic nitrogen dynamics in the Crocodile River (South Africa). Water, Air, and Soil Pollution **155**:303-320.

Delleur, J. W. 2003. The evolution of urban hydrology: past, present, and future. Journal of hydraulic engineering **129**:563-573.

Demšar, J. 2006. Statistical comparisons of classifiers over multiple data sets. Journal of Machine learning research **7**:1-30.

Dietz, M. E. 2007. Low impact development practices: A review of current research and recommendations for future directions. Water, Air, and Soil Pollution **186**:351-363.

Doeg, T., and G. Milledge. 1991. Effect of experimentally increasing concentration of suspended sediment on macroinvertebrate drift. Marine and Freshwater Research **42**:519-526.

Dohner, E., A. Markowitz, M. Barbour, J. Simpson, J. Byrne, and G. Dates. 1997. Volunteer Stream Monitoring: A Methods Manual. Environmental Protection Agency, Washington, DC, USA.

Dominguez-Granda, L., K. Lock, and P. L. Goethals. 2011. Using multi-target clustering trees as a tool to predict biological water quality indices based on benthic macroinvertebrates and environmental parameters in the Chaguana watershed (Ecuador). Ecological Informatics **6**:303-308.

Domisch, S., S. C. JAeHNIG, and P. Haase. 2011. Climate-change winners and losers: stream macroinvertebrates of a submontane region in Central Europe. Freshwater Biology **56**:2009-2020.

Douglas-Mankin, K., R. Srinivasan, and J. Arnold. 2010. Soil and Water Assessment Tool (SWAT) model: Current developments and applications. Transactions of the ASABE **53**:1423-1431.

Durazno, G. 2013. Variacion de flujo y de cargas contaminantes en el afluente y cargas de salida de la planta de aguas residuales de Ucubamba - año 2012. Empresa Municipal de Telecomunicaciones, Agua Potable, Alcantarillado y Saneamiento, Cuenca Ecuador.

Džeroski, S., D. Demšar, and J. Grbović. 2000. Predicting chemical parameters of river water quality from bioindicator data. Applied Intelligence **13**:7-17.

Džeroski, S., J. Grbović, W. J. Walley, and B. Kompare. 1997. Using machine learning techniques in the construction of models. II. Data analysis with rule induction. Ecological Modelling **95**:95-111.

Encalada, A. C., M. R. Sant, and N. Prat i Fornells. 2011. Protocolo simplificado y guía de evaluación de la calidad ecológica de ríos andinos (CERA-S). Proyecto FUCARA, Quito - Ecuador.

EPA. 1999. Combined sewer overflows–Guidance for monitoring and modeling. Environmental Protection Agency.

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

Espinoza Berrezueta, D. A., and T. C. Zumba López. 2018. Estudio del impacto ambiental en la calidad del agua del río Cuenca producido por la descarga del efluente de la planta de tratamiento de aguas residuales de Ucubamba. Universidad de Cuenca, Cuenca Ecuador.

Esquivel, J. C., B. Verbeiren, A. Alvarado, J. Feyen, and F. Cisneros. 2008. Preliminary statistical analysis of the water quality database of ETAPA. . PROMAS - Universidad de Cuenca, Cuenca - Ecuador.

Estrella, R., and V. Tobar. 2013. Hidrologia y Climatologia - Formulación del plan de manejo integral de la subcuenca del río Machangara. ACOTECNIC Cia. Ltda. - Consejo de Cuenca del Rio Machangara, Cuenca - Ecuador.

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. 2015. Flow rating curves of the Cuenca River gauging stations. ETAPA-EP.

ETAPA-EP. 2017. Weekly records of the water quality of the Ucumaba wastewater treatment plant of the years 2015 and 2016. ETAPA-EP, Ecuador.

Everaert, G., P. Boets, K. Lock, S. Džeroski, and P. L. Goethals. 2011. Using classification trees to analyze the impact of exotic species on the ecological assessment of polder lakes in Flanders, Belgium. Ecological Modelling **222**:2202-2212.

Everaert, G., J. De Neve, P. Boets, L. Dominguez-Granda, S. T. Mereta, A. Ambelu, T. H. Hoang, P. L. Goethals, and O. Thas. 2014. Comparison of the abiotic preferences of macroinvertebrates in tropical river basins.

Everaert, G., I. S. Pauwels, P. Boets, F. Buysschaert, and P. L. Goethals. 2013a. Development and assessment of ecological models in the context of the European Water Framework Directive: Key issues for trainers in data-driven modeling approaches. Ecological Informatics **17**:111-116.

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

FAO. 2016. Ecuador. AQUASTAT website. Food and Agriculture Organization of the United Nations (FAO). Website accessed on [2019/01/04].

Feldman, R. S., and E. F. Connor. 1992. The relationship between pH and community structure of invertebrates in streams of the Shenandoah National Park, Virginia, USA. Freshwater Biology **27**:261-276.

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

Fewtrell, L., J. Bartram, and W. W. H. Organization. 2001. Water quality: Guidelines, standards, and health: Assessment of risk and risk management for water-related infectious disease. IWA Publishing, London , UK.

Folks, J. L., and R. S. Chhikara. 1978. The inverse Gaussian distribution and its statistical application-a review. Journal of the Royal Statistical Society. Series B (Methodological):263-289.

Forbes, C., M. Evans, N. Hastings, and B. Peacock. 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., P. L. Goethals, K. Lock, V. Asio, M. Bande, and O. Thas. 2018. Model-based analysis of the relationship between macroinvertebrate traits and environmental river conditions. Environmental Modelling & Software **106**:57-67.

Forio, M. A. E., D. Landuyt, E. Bennetsen, K. Lock, T. H. T. Nguyen, M. N. D. Ambarita, P. L. S. Musonge, P. Boets, G. Everaert, L. Dominguez-Granda, and P. L. M. Goethals. 2015. Bayesian belief network models to analyse and predict ecological water quality in rivers. Ecological Modelling **312**:222-238.

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

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

Friberg, N., J. Skriver, S. E. LARSEN, M. L. Pedersen, and A. Buffagni. 2010. Stream macroinvertebrate occurrence along gradients in organic pollution and eutrophication. Freshwater Biology **55**:1405-1419.

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

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

Gabriels, W., K. Lock, N. De Pauw, and P. L. Goethals. 2010. Multimetric Macroinvertebrate Index Flanders (MMIF) for biological assessment of rivers and lakes in Flanders (Belgium). Limnologica-Ecology and Management of Inland Waters **40**:199-207.

Galay, V. 1983. Causes of river bed degradation. Water Resources Research **19**:1057-1090.

Goethals, P. 2005. Data driven development of predictive ecological models for benthic macroinvertebrates in rivers. Ghent University, Ghent - Belgium.

Goethals, P., A. Dedecker, W. Gabriëls, and N. De Pauw. 2006. Development and application of predictive river ecosystem models based on classification trees and artificial neural networks. Pages 151-167 Ecological Informatics. Springer, Berlin, Heidelberg.

Goethals, P., and M. Forio. 2018. Advances in Ecological Water System Modeling: Integration and Leanification as a Basis for Application in Environmental Management. Multidisciplinary Digital Publishing Institute.

Goethals, P. L., A. P. Dedecker, W. Gabriels, S. Lek, and N. De Pauw. 2007. Applications of artificial neural networks predicting macroinvertebrates in freshwaters. Aquatic Ecology **41**:491-508.

Gofti-Laroche, L., D. Demanse, J.-C. Joret, and D. Zmirou. 2003. Health risks and parasitical quality of water. Journal (American Water Works Association) **95**:162-172.

Gostner, W., M. Alp, A. J. Schleiss, and C. T. Robinson. 2013. The hydro-morphological index of diversity: a tool for describing habitat heterogeneity in river engineering projects. Hydrobiologia **712**:43-60.

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

Greeley & Hansen, and ACSAM. 2017. Diseños definitivos de la planta de tratamiento de aguas residuales de Guangarcucho. ETAPA-EP, Cuenca - Ecuador.

Griffiths, M. 2002. The European Water Framework Directive: an approach to integrated river basin management. European Water Management Online **5**:1-14.

Guisan, A., A. Lehmann, S. Ferrier, M. Austin, J. OVERTON, R. Aspinall, and T. Hastie. 2006. Making better biogeographical predictions of species' distributions. Journal of Applied Ecology **43**:386-392.

Haaland, S., D. Hongve, H. Laudon, G. Riise, and R. Vogt. 2010. Quantifying the drivers of the increasing colored organic matter in boreal surface waters. Environmental science & technology **44**:2975-2980.

Hardin, J. W., and J. M. Hilbe. 2013. Generalized Estimating Equations Second Edition edition. CRC Press

Taylor & Francis Group, Boca Raton, FL 33487-2742.

Harremoës, P. 1982. Immediate and delayed oxygen depletion in rivers. Water Research **16**:1093-1098.

Hawkins, C. P., J. N. Hogue, L. M. Decker, and J. W. Feminella. 1997. Channel morphology, water temperature, and assemblage structure of stream insects. Journal of the North American Benthological Society **16**:728-749.

Heino, J., H. Mykrä, H. Hämäläinen, J. Aroviita, and T. Muotka. 2007. Responses of taxonomic distinctness and species diversity indices to anthropogenic impacts and natural environmental gradients in stream macroinvertebrates. Freshwater Biology **52**:1846-1861.

Helson, J. E., and D. D. Williams. 2013. Development of a macroinvertebrate multimetric index for the assessment of low-land streams in the neotropics. Ecological Indicators **29**:167-178.

Henze, M., W. Gujer, T. Mino, and M. Van Loosdrecht. 2000. Activated sludge models ASM1, ASM2, ASM2d and ASM3. IWA publishing.

Hering, D., C. K. Feld, O. Moog, and T. Ofenböck. 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.

Hoang, T. H., K. Lock, K. C. Dang, N. De Pauw, and P. Goethals. 2010a. Spatial and temporal patterns of macroinvertebrate communities in the Du River basin in northern Vietnam. Journal of Freshwater Ecology **25**:637-647.

Hoang, T. H., K. Lock, A. Mouton, and P. L. Goethals. 2010b. Application of classification trees and support vector machines to model the presence of macroinvertebrates in rivers in Vietnam. Ecological Informatics **5**:140-146.

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

Holguin-Gonzalez, J. E., P. Boets, G. Everaert, I. S. Pauwels, K. Lock, S. Gobeyn, L. Benedetti, Y. Amerlinck, I. Nopens, and P. L. M. Goethals. 2014. Development and assessment of an integrated ecological modelling framework to assess the effect of investments in wastewater treatment on water quality. Water Science and Technology **70**:1798-1807.

Holguin-Gonzalez, J. E., G. Everaert, P. Boets, A. Galvis, and P. L. M. Goethals. 2013b. Development and application of an integrated ecological modelling framework to analyze the impact of wastewater discharges on the ecological water quality of rivers. Environmental Modelling & Software **48**:27-36.

Holomuzki, J. R., and B. J. Biggs. 2003. Sediment texture mediates high-flow effects on lotic macroinvertebrates. Journal of the North American Benthological Society **22**:542-553.

Holzer, P., and P. Krebs. 1998. Modelling the total ammonia impact of CSO and WWTP effluent on the receiving water. Water Science and Technology **38**:31-39.

Hu, B., J. Shao, and M. Palta. 2006. Pseudo-R 2 in logistic regression model. Statistica Sinica:847-860.

Hvitved-Jacobsen, T. 1982. The impact of combined sewer overflows on the dissolved oxygen concentration of a river. Water Research **16**:1099-1105.

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. 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. Archiv für Hydrobiologie **143**:179-195.

Jacobsen, D. 2005. Temporally variable macroinvertebrate–stone relationships in streams. Hydrobiologia **544**:201-214.

Jacobsen, D. 2008. Tropical high-altitude streams. Tropical stream ecology:219-256.

Jacobsen, D., and A. Encalada. 1998. The macroinvertebrate fauna of Ecuadorian highland streams in the wet and dry season. Archiv für Hydrobiologie **142**:53-70.

Jacobsen, D., and R. Marín. 2008. Bolivian Altiplano streams with low richness of macroinvertebrates and large diel fluctuations in temperature and dissolved oxygen. Aquatic Ecology **42**:643-656.

Jacobsen, D., S. Rostgaard, and J. J. Vásconez. 2003. Are macroinvertebrates in high altitude streams affected by oxygen deficiency? Freshwater Biology **48**:2025-2032.

Jerves-Cobo, R., G. Córdova-Vela, X. Iñiguez-Vela, C. Díaz-Granda, W. Van Echelpoel, F. Cisneros, I. Nopens, and P. Goethals. 2018a. Model-Based Analysis of the Potential of Macroinvertebrates as Indicators for Microbial Pathogens in Rivers. Water **10**:375.

Jerves-Cobo, R., G. Everaert, X. Iñiguez-Vela, G. Córdova-Vela, C. Díaz-Granda, F. Cisneros, I. Nopens, and P. L. Goethals. 2017. A Methodology to Model Environmental Preferences of EPT Taxa in the Machangara River Basin (Ecuador). Water **9**:195.

Jerves-Cobo, R., M. A. E. Forio, K. Lock, J. Van Butsel, G. Pauta, F. Cisneros, I. Nopens, and P. Goethals. 2019. Biological water quality in tropical rivers in dry and rainy seasons: A model-based analysis. Submitted for publication.

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

Jorgensen, S. E., S. N. Nielsen, and L. A. Jorgensen. 1991. Handbook of Ecological Parameters and Ecotoxicology. Elsevier.

Junqueira, V., and S. Campos. 1998. Adaptation of the “BMWP” method for water quality evaluation to Rio das Velhas watershed (Minas Gerais, Brazil). Acta Limnologica Brasiliensia **10**:125-135.

Kasangaki, A., L. J. Chapman, and J. Balirwa. 2008. Land use and the ecology of benthic macroinvertebrate assemblages of high-altitude rainforest streams in Uganda. Freshwater Biology **53**:681-697.

Kauffman, J. B., and W. C. Krueger. 1984. Livestock impacts on riparian ecosystems and streamside management implications... a review. Journal of range management **37**:430-438.

Kay, D., and A. McDonald. 1983. Predicting coliform concentrations in upland impoundments: design and calibration of a multivariate model. Applied and environmental microbiology **46**:611-618.

Kelso, B. H., R. V. Smith, and R. J. Laughlin. 1999. Effects of carbon substrates on nitrite accumulation in freshwater sediments. Applied and environmental microbiology **65**:61-66.

Kohavi, R., and F. Provost. 1998. Glossary of terms. Machine Learning **30**:271-274.

Krejci, V., W. Schilling, and S. Gammeter. 1994. Receiving water protection during wet weather. Water Science and Technology **29**:219-229.

Lam, D., L. Leon, S. Hamilton, N. Crookshank, D. Bonin, and D. Swayne. 2004. Multi-model integration in a decision support system: a technical user interface approach for watershed and lake management scenarios. Environmental Modelling & Software **19**:317-324.

Landis, J. R., and G. G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. Biometrics **33**:159-174.

Leclerc, H., D. Mossel, S. Edberg, and C. Struijk. 2001. Advances in the bacteriology of the coliform group: their suitability as markers of microbial water safety. Annual Reviews in Microbiology **55**:201-234.

Lek, S. 2007. Uncertainty in ecological models. Ecological Modelling **1**:1-2.

Lenat, D. R., and V. H. Resh. 2001. Taxonomy and stream ecology—the benefits of genus-and species-level identifications. Journal of the North American Benthological Society **20**:287-298.

Lior, R. 2014. Data mining with decision trees: theory and applications. World Scientific, Singapore.

Liu, Y., and H. V. Gupta. 2007. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. Water Resources Research **43**.

Lock, K., and P. L. Goethals. 2008. Distribution and ecology of the stoneflies (Plecoptera) of Flanders (Belgium). Pages 203-213 *in* Annales de Limnologie-International Journal of Limnology. Cambridge Univ Press.

Lock, K., and P. L. Goethals. 2012. Distribution and ecology of the caddisflies (Trichoptera) of Flanders (Belgium). Pages 31-37 *in* Annales de Limnologie-International Journal of Limnology. Cambridge Univ Press.

Lock, K., and P. L. Goethals. 2013. Habitat suitability modelling for mayflies (Ephemeroptera) in Flanders (Belgium). Ecological Informatics **17**:30-35.

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

Mac Berthouex, P., and L. C. Brown. 2002. Statistics for environmental engineers. Lewis publishers Boca Raton, USA.

Mac Nally, R. 2000. Regression and model-building in conservation biology, biogeography and ecology: the distinction between–and reconciliation of–'predictive'and 'explanatory'models. Biodiversity & Conservation **9**:655-671.

MAE-Ecuador. 2015. TULAS - Texto Unificado de Legislación Secundaria.*in* Ministerio-del-Ambiente, editor. Registro Oficial, Quito - Ecuador.

Mahloch, J. L. 1974. Comparative analysis of modeling techniques for coliform organisms in streams. Applied microbiology **27**:340-345.

Maillard, P., and N. A. P. Santos. 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. Journal of environmental management **86**:158-170.

Maimon, O., and L. Rokach. 2005. Data mining and knowledge discovery handbook. Springer, New York - USA.

Mallin, M. A., V. L. Johnson, S. H. Ensign, and T. A. MacPherson. 2006. Factors contributing to hypoxia in rivers, lakes, and streams. Limnology and Oceanography **51**:690-701.

Manel, S., H. C. Williams, and S. J. Ormerod. 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. Journal of Applied Ecology **38**:921-931.

Marsalek, J., M. Karamouz, B. J. Cisneros, P.-A. Malmquist, J. A. Goldenfum, and B. Chocat. 2014. Urban Water Cycle Processes and Interactions: Urban Water Series-UNESCO-IHP. CRC Press.

McCarty, P. L., and C. Brodersen. 1962. Theory of extended aeration activated sludge. Journal (Water Pollution Control Federation):1095-1103.

McCullagh, P. 1984. Generalized linear models. European Journal of Operational Research **16**:285-292.

McDowell, L. R. 1996. Feeding minerals to cattle on pasture. Animal feed science and technology **60**:247-271.

Meirlaen, J. 2002. Immission based real-time control of the integrated urban wastewater system. Ghent University.

Meirlaen, J., B. Huyghebaert, F. Sforzi, L. Benedetti, and P. Vanrolleghem. 2001. Fast, simultaneous simulation of the integrated urban wastewater system using mechanistic surrogate models. Water Science and Technology **43**:301-309.

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

Meybeck, M., L. Laroche, H. Dürr, and J. Syvitski. 2003. Global variability of daily total suspended solids and their fluxes in rivers. Global and planetary change **39**:65-93.

Moisen, G. 2008. Classification and regression trees. Encyclopedia of Ecology **1**:582-588.

Montalvo, C., R. Jerves, and L. Domínguez. 2018. Determinación de cargas de contaminación en aliviaderos de la red de alcantarillado combinado de la ciudad de Cuenca - Ecuador. Page 9 XXXVI Congreso Interamericano de Ingeniería Sanitaria y Ambiental. Asociación Interamericana de Ingeniería Sanitaria y Ambiental – AIDIS, Guayaquil - Ecuador.

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. Internationale Vereinigung für Theoretische und Angewandte Limnologie: Mitteilungen **22**:1-23.

Mouton, A. M., B. De Baets, and P. L. Goethals. 2009a. Knowledge-based versus data-driven fuzzy habitat suitability models for river management. Environmental Modelling & Software **24**:982-993.

Mouton, A. M., H. Van Der Most, A. Jeuken, P. L. Goethals, and N. De Pauw. 2009b. Evaluation of river basin restoration options by the application of the water framework directive explorer in the Zwalm River basin (Flanders, Belgium). River research and applications **25**:82-97.

Moya, N., R. M. Hughes, E. Domínguez, F.-M. Gibon, E. Goitia, and T. Oberdorff. 2011. Macroinvertebrate-based multimetric predictive models for evaluating the human impact on biotic condition of Bolivian streams. Ecological Indicators **11**:840-847.

Mulliss, R., D. Revitt, and R. Shutes. 1997. The impacts of discharges from two combined sewer overflows on the water quality of an urban watercourse. Water Science and Technology **36**:195-199.

Mustapha, A., A. Z. Aris, M. F. Ramli, and H. Juahir. 2012. Spatial-temporal variation of surface water quality in the downstream region of the Jakara River, north-western Nigeria: A statistical approach. Journal of Environmental Science and Health, Part A **47**:1551-1560.

Mustow, S. 2002. Biological monitoring of rivers in Thailand: use and adaptation of the BMWP score. Hydrobiologia **479**:191-229.

Najafpour, S., A. Alkarkhi, M. Kadir, and G. D. Najafpour. 2008. Evaluation of Spatial and Temporal Variation in River Water Quality. International Journal of Environmental Research **2**.

Nguyen, H. H., G. Everaert, W. Gabriels, T. H. Hoang, and P. L. Goethals. 2014. A multimetric macroinvertebrate index for assessing the water quality of the Cau river basin in Vietnam. Limnologica-Ecology and Management of Inland Waters **45**:16-23.

Nguyen, T. H. T., P. Boets, K. Lock, M. N. D. Ambarita, M. A. E. Forio, P. Sasha, L. E. Dominguez-Granda, T. H. T. Hoang, G. Everaert, and P. L. Goethals. 2015. Habitat suitability of the invasive water hyacinth and its relation to water quality and macroinvertebrate diversity in a tropical reservoir. Limnologica-Ecology and Management of Inland Waters **52**:67-74.

NYC Environmental Protection. 2012. Criteria for detention facility design. Page 17 *in* B. o. W. S. Operations, editor., NY.

Ogleni, N., and B. Topal. 2011. Water quality assessment of the Mudurnu River, Turkey, using biotic indices. Water resources management **25**:2487.

Oliver, B. G. 1984. Guidelines for drining-water quality, volume 1: Recommendations, . Elsevier, Geneva - Switzerland.

Palmer, M. A., E. Bernhardt, J. Allan, P. Lake, G. Alexander, S. Brooks, J. Carr, S. Clayton, C. Dahm, and J. F. Shah. 2005. Standards for ecologically successful river restoration. Journal of Applied Ecology **42**:208-217.

Parkyn, S. M., R. J. Davies-Colley, A. B. Cooper, and M. J. Stroud. 2005. Predictions of stream nutrient and sediment yield changes following restoration of forested riparian buffers. Ecological Engineering **24**:551-558.

Parsons, M., M. Thoms, and R. Norris. 2002. Australian river assessment system: AusRivAS physical assessment protocol. Monitoring river health initiative technical report.

Passerat, J., N. K. Ouattara, J.-M. Mouchel, V. Rocher, and P. Servais. 2011. Impact of an intense combined sewer overflow event on the microbiological water quality of the Seine River. Water Research **45**:893-903.

Pearce, J., and S. Ferrier. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling **133**:225-245.

Philips, S., H. J. Laanbroek, and W. Verstraete. 2002. Origin, causes and effects of increased nitrite concentrations in aquatic environments. Reviews in environmental science and biotechnology **1**:115-141.

PROMAS-UCuenca. 2005. Estudio fluvio-morfológico del río Tomebamba. Cuenca, Ecuador.

PROMAS-UCuenca. 2010. Información de la red meteorológica e hidrológica. Progarma para el Manejo del Agua y el Suelo. Universidad de Cuenca, Cuenca - Ecuador.

Quinlan, J. R. 1986. Induction of decision trees. Machine Learning **1**:81-106.

Quinlan, J. R. 1987. Generating production rules from decision trees. Pages 304-307 *in* 10th International Joint Conference on Artificial Intelligence. Morgan Kaufmann Publishers Inc., Milan, Italy.

Quinlan, J. R. 2014. C4. 5: programs for machine learning. Elsevier, San Mateo, CA, USA.

R-Core-Team. 2015. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. R Foundation for Statistical Computing.

R-Core-Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2016.

Rahaman, M. M., and O. Varis. 2005. Integrated water resources management: evolution, prospects and future challenges. Sustainability: science, practice and policy **1**:15-21.

Rantz, S. E. 1982. Measurement and Computation of Streamflow: volume 2, computation of discharge. USGPO.

Rauch, W., M. Henze, L. Koncsos, P. Reichert, P. Shanahan, L. Somlyódy, and P. Vanrolleghem. 1998. River water quality modelling: I. State of the art. Water Science and Technology **38**:237-244.

Raven, P. J. 1998. River Habitat Quality: the physical character of rivers and streams in the UK and Isle of Man. Environment Agency.

Reichert, P., D. Borchardt, M. Henze, W. Rauch, P. Shanahan, L. Somlyódy, and P. Vanrolleghem. 2001. River Water Quality Model No. 1. IWA Scientific and Technical Report No. 12. IWA Publishing, London, UK. ISBN 1-900222-82, 2001, 131.

Ríos-Touma, B., R. Acosta, and N. Prat. 2014. The Andean Biotic Index (ABI): revised tolerance to pollution values for macroinvertebrate families and index performance evaluation. Revista de Biología Tropical **62**:249-273.

Ríos-Touma, B., A. C. Encalada, and N. Prat Fornells. 2011. Macroinvertebrate Assemblages of an Andean High-Altitude Tropical Stream: The Importance of Season and Flow. International review of hydrobiology **96**:667-685.

Roldán Pérez, G. 1999. Los macroinvertebrados y su valor como indicadores de la calidad del agua. Academia Colombiana de Ciencia **23**:375-387.

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.

Rondinini, C., M. Di Marco, F. Chiozza, G. Santulli, D. Baisero, P. Visconti, M. Hoffmann, J. Schipper, S. N. Stuart, and M. F. Tognelli. 2011. Global habitat suitability models of terrestrial mammals. Philosophical Transactions of the Royal Society of London B: Biological Sciences **366**:2633-2641.

Rosenberg, D. M., and V. H. Resh. 1993. Freshwater biomonitoring and benthic macroinvertebrates. Chapman & Hall.

Rossman, L., and W. Huber. 2016. Storm Water Management Model Reference Manual Volume I–Hydrology (Revised). US Environmental Protection Agency: Cincinnati, OH, USA.

Rousseeuw, P. J., and A. M. Leroy. 2005. Robust regression and outlier detection. John Wiley & Sons.

Rubio, R., D. Mora, R. Jerves, and M. Arias-Hidalgo. 2017. Estimación de las pérdidas de la precipitación en una cuenca hidrográfica urbana de la Ciudad de Cuenca. Page 4 First international Conference on Integrated Water Quality Management. Universidad de Cuenca, Cuenca - Ecuador.

Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto. 2004. Sensitivity analysis in practice: a guide to assessing scientific models. Chichester, England.

Sánchez, L., A. Sánchez, G. Galvis, and J. Latorre. 2007. Filtración en múltiples etapas. IRC International Water and Sanitation Centre, Delf - The Netherlands.

Schleiter, I. M., D. Borchardt, R. Wagner, T. Dapper, K.-D. Schmidt, H.-H. Schmidt, and H. Werner. 1999. Modelling water quality, bioindication and population dynamics in lotic ecosystems using neural networks. Ecological Modelling **120**:271-286.

Schrijver, L. D. 2017. Determination of the water quality of the sewage overflows from Cuenca and the impact on the receiving surface water. Hogeschool Gent, Ghent - Belgium.

Schwarzenbach, R. P., T. Egli, T. B. Hofstetter, U. von Gunten, and B. Wehrli. 2010. Global water pollution and human health. Annual Review of Environment and Resources **35**:109-136.

SENPLADES. 2016. Proyecciones referenciales de población a nivel de distritos de planificación: 2010-2020. Secretaría Nacional de Planificación y Desarrollo Ecuador, Subsecretaría de Información, Dirección de Normas y Metodología, Quito - Ecuador.

Seyfried, P., and E. Harris. 1990. Bacteriological characterization of feces and source differentiation. Water Resources Branch, Ontario Ministry of the Environment, Toronto, Ontario, Canada.

Sharp, W. 1971. A Topologically Optimum Water-Sampling Plan for Rivers and Streams. Water Resources Research **7**:1641-1646.

Shmueli, G. 2010. To explain or to predict? Statistical Science **25**:289-310.

Solvi, A.-M. 2006. Modelling the sewer-treatment-urban river system in view of the EU Water Framework Directive. PhD thesis, Ghent University, Belgium.

Solvi, A., L. Benedetti, S. Gillé, P. Schosseler, A. Weidenhoupt, and P. Vanrolleghem. 2005. Integrated urban catchment modelling for a sewer-treatment-river system.*in* 10th International Conference on Urban Drainage.

Stevens, J. P. 1984. Outliers and influential data points in regression analysis. Psychological Bulletin **95**:334.

Stockwell, D. R., and A. T. Peterson. 2002. Effects of sample size on accuracy of species distribution models. Ecological Modelling **148**:1-13.

Strobl, R. O., and P. D. Robillard. 2008. Network design for water quality monitoring of surface freshwaters: A review. Journal of environmental management **87**:639-648.

Stutter, M., S. Richards, and J. Dawson. 2013. Biodegradability of natural dissolved organic matter collected from a UK moorland stream. Water Research **47**:1169-1180.

Tallon, P., B. Magajna, C. Lofranco, and K. T. Leung. 2005. Microbial indicators of faecal contamination in water: a current perspective. Water, Air, and Soil Pollution **166**:139-166.

Te Chow, V. 1959. Open-channel hydraulics. McGraw-Hill New York.

Thode, H. C. 2002. Testing for normality. CRC press.

Thorne, R., and P. Williams. 1997. The response of benthic macroinvertebrates to pollution in developing countries: a multimetric system of bioassessment. Freshwater Biology **37**:671-686.

Thuiller, W. 2003. BIOMOD–optimizing predictions of species distributions and projecting potential future shifts under global change. Global change biology **9**:1353-1362.

Ting, K. M. 2002. An instance-weighting method to induce cost-sensitive trees. Ieee Transactions on Knowledge and Data Engineering **14**:659-665.

Tomsic, C. A., T. C. Granata, R. P. Murphy, and C. J. Livchak. 2007. Using a coupled eco-hydrodynamic model to predict habitat for target species following dam removal. Ecological Engineering **30**:215-230.

UN-Water. 2011. Policy brief: water quality. Policy and Analytical Briefs:22.

UN-WWAP, U. N. W. W. A. P. 2017. The United Nations World Water Development Report 2017: The Untapped Resource. UNESCO, Paris.

USEPA - U.S. Environmental Protection Agency. 2011. A Primer on Using Biological Assessments to Support Water Quality Management. Washington, DC.

USEPA. 1997. Monitoring, Volunteer Stream: A Methods Manual. US Environmental Protection Agency, Office of Wetlands, Oceans, and Watersheds, Washington, DC.

Van Horne, B., and J. A. Wiens. 1991. Forest bird habitat suitability models and the development of general habitat models. Colorado State University Fort Collins.

Vanhooren, H., J. Meirlaen, Y. Amerlinck, F. Claeys, H. Vangheluwe, and P. A. Vanrolleghem. 2003. WEST: modelling biological wastewater treatment. Journal of Hydroinformatics **5**:27-50.

Vanrolleghem, P., and S. Gillot. 2002. Robustness and economic measures as control benchmark performance criteria. Water Science and Technology **45**:117-126.

Vaughan, I. P., and S. J. Ormerod. 2005. The continuing challenges of testing species distribution models. Journal of Applied Ecology **42**:720-730.

Vayssières, M. P., R. E. Plant, and B. H. Allen-Diaz. 2000. Classification trees: An alternative non-parametric approach for predicting species distributions. Journal of vegetation science **11**:679-694.

Venables, W. N., and B. D. Ripley. 2002. Tree-based methods. Pages 251-269 Modern Applied Statistics with S. Springer, New York.

Wade, T. J., R. L. Calderon, E. Sams, M. Beach, K. P. Brenner, A. H. Williams, and A. P. Dufour. 2006. Rapidly Measured Indicators of Recreational Water Quality Are Predictive of Swimming-Associated Gastrointestinal Illness. Environmental Health Perspectives **114**:24-28.

Walsh, C. J., A. H. Roy, J. W. Feminella, P. D. Cottingham, P. M. Groffman, and R. P. Morgan. 2005. The urban stream syndrome: current knowledge and the search for a cure. Journal of the North American Benthological Society **24**:706-723.

Ward, J. V., and J. Stanford. 1995. Ecological connectivity in alluvial river ecosystems and its disruption by flow regulation. Regulated rivers: research & management **11**:105-119.

Ward, R. C. 1973. Data acquisition systems in water quality management.

Wei, J.-M., X.-J. Yuan, Q.-H. Hu, and S.-Q. Wang. 2010. A novel measure for evaluating classifiers. Expert Systems with Applications **37**:3799-3809.

Weyrauch, P., A. Matzinger, E. Pawlowsky-Reusing, S. Plume, D. von Seggern, B. Heinzmann, K. Schroeder, and P. Rouault. 2010. Contribution of combined sewer overflows to trace contaminant loads in urban streams. Water Research **44**:4451-4462.

White, J. D., and J. A. Dracup. 1977. Water quality modeling of a high mountain stream. Journal (Water Pollution Control Federation):2179-2189.

WHO. 2003. Guidelines for safe recreational water environments: Coastal and fresh waters. World Health Organization, Geneva - Switzerland.

WHO. 2004. Guidelines for drinking-water quality. World Health Organization, Geneva - Switzerland.

Wilcock, R. J., R. Battino, W. F. Danforth, and E. Wilhelm. 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. The Journal of Chemical Thermodynamics **10**:817-822.

Wilkinson, J., A. Jenkins, M. Wyer, and D. Kay. 1995. Modelling faecal coliform dynamics in streams and rivers. Water Research **29**:847-855.

Wilson, A., R. Watts, and M. Stevens. 2008. Effects of different management regimes on aquatic macroinvertebrate diversity in Australian rice fields. Ecological Research **23**:565-572.

Witten, I. H., and E. Frank. 2005. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, San Francisco - USA.

Yu, R., and M. Abdel-Aty. 2013. Utilizing support vector machine in real-time crash risk evaluation. Accident Analysis & Prevention **51**:252-259.

Zabel, T., I. Milne, and G. Mckay. 2001. Approaches adopted by the European Union and selected Member States for the control of urban pollution. Urban Water **3**:25-32.

Zarkami, R. 2008. Habitat suitability modelling of pike (Esox lucius) in rivers. Ghent University.

Zúñiga, M. d. C., W. Cardona, J. Cantera, Y. Carvajal, and L. Castro. 2009. Bioindicadores de calidad de agua y caudal ambiental. Caudal ambiental: Conceptos Experiencias y Desafíos. Universidad del Valle: Cali, Colombia **1**:303-310.

Zuur, A., E. N. Ieno, N. Walker, A. A. Saveliev, and G. M. Smith. 2009. Mixed effects models and extensions in ecology with R. Springer Science & Business Media.

CURRICULUM VITAE

Personal information

Family names:	Jerves Cobo
First names:	Rubén Fernando
Birth date:	December 4th, 1969
Place of birth:	Cuenca, Ecuador
Gender:	Male
Civil state:	Married
e-mail:	rfjerves@yahoo.com
	rubenf.jervesc@ucuenca.edu.ec
ORCID:	0000-0002-7141-2390
ResearcherID**:**	G-4062-2018
Scopus Author ID:	57193561638
Publons ID:	1352013

Education

2015 – 2019 Doctoral studies in Applied Biological Sciences
Ghent University
Ghent, Belgium
Thesis: Integrated water system modelling to support water management in the Cuenca River basin.

1999-2000 Master's Degree in Engineering and Environmental Management
Polytechnic University of Catalonia
Barcelona, Spain
Thesis: Air Emissions Modelling due to an industrial activity. Power Station El Descanso, Azuay Province, Republic of Ecuador.

1987-1993 Civil Engineering
Cuenca University
Cuenca, Ecuador
Thesis: Integrated water supply system for the Checa Parish, Azuay Province, Republic of Ecuador.

Training courses

Feb-May 2018 **Advanced Academic English: Writing Skills (Bioscience) Engineering**. UCT, Ghent University. (3 crdts)

Feb-Jun 2017 **Natural Systems for (Waste)Water Treatment**. Faculty of Bioscience Engineering, Campus Coupure, Ghent University. (3 crdts)

May 2017	**R Intermediate Course.** Flanders Training Network for Methodology and Statistics (FLAMES). Campus De Sterre, Ghent University. (14 hrs)
Feb 2017	**Introduction to Categorical Data Analysis with R.** Flanders Training Network for Methodology and Statistics (FLAMES). Campus Dunant, Ghent University. (18 hrs)
Sep 2016 to Jan 2017	**Integrated Modelling and Design and Basin Management Plans**. Faculty of Bioscience Engineering, Campus Coupure, Ghent University. (3 crdts)
Sep 2016 to Jan 2017	**Water Quality Modelling.** Faculty of Bioscience Engineering, Campus Coupure, Ghent University. (3 crdts)
Sep 2015 to Jan 2016	**Modelling and Control of Waste Water Treatment Plants**. Faculty of Bioscience Engineering, Campus Coupure, Ghent University. (3 crdts)
Oct 2016	**Introduction to R**. Flanders Training Network for Methodology and Statistics (FLAMES). Campus De Sterre, Ghent University. (14 hrs)
Jun 2016	**Leadership Foundation Course,** Ghent University. (25.5 hrs)
Oct-Dec 2015	**Advanced Academic English: Conference Skills**. UCT, Ghent University. (2 crdts)
Oct 2015	**Project management**. Ghent University. (24 hrs)
Feb-Jun 2015	**Water Quality Management**. Faculty of Bioscience Engineering, Campus Coupure, Ghent University. (5 crdts)
Sep-Dec 2015	**Practical English 4 (B1).** UCT, Ghent University. (3 crdts)
Sep-Dec 2016	**Practical English 5 (B2).** UCT, Ghent University. (3 crdts)
Oct-Dec 2016	**Upper-intermediate Academic English (B2).** UCT, Ghent University. (3 crdts)
Sep-Nov 2017	**Lecturing skills in English**. UCT, Ghent University. (16 hrs)

Personal profile

My personal experience is focused on the academic, scientific, research fields, mainly on topics related to integrated water system modelling for decision support in rivers' management. I have professional experience in consulting on projects associated with water supply systems, urban sewage systems, wastewater treatment and air pollution management.

Organization

	Position	*Date*
International Water Association (IWA)	member	2015 to present
Ecuadorian Association of Sanitary and Environmental Engineering (AEISA)	member	2001 to present
Ecuadorian Association of Civil Engineers (CICE)	member	1993 to present
Environmental Management Commission of Cuenca, Ecuador (CGA)	Board member	2004-2005

Publications

A1 peer-reviewed papers

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

Jerves-Cobo, R.; Córdova-Vela, G.; Iñiguez-Vela, X.; Díaz-Granda, C.; Van Echelpoel, W.; Cisneros, F.; Nopens, I.; Goethals, P.L.M. 2018. Model-Based Analysis of the Potential of Macroinvertebrates as Indicators for Microbial Pathogens in Rivers. Water, vol. *10*, p. 375.

Jerves-Cobo, R., G. Everaert, X. Iñiguez-Vela, G. Córdova-Vela, C. Díaz-Granda, F. Cisneros, I. Nopens, P.L.M. Goethals. 2017. A Methodology to Model Environmental Preferences of EPT Taxa in the Machangara River Basin (Ecuador). Water, vol. *9*, p. 195.

A1 papers (under review and in preparation)

Jerves-Cobo R, Lock K, Van Butsel J, Pauta G, Cisneros F, Nopens I, Goethals PLM, Integrated ecological modelling for evidence-based determination of water management interventions in urbanized river basins: case study in the Cuenca River basin (Ecuador) (Under review).

Jerves-Cobo R, Forio M. A. E., Lock K, Van Butsel J, Pauta G, Cisneros F, Nopens I, Goethals PLM, Biological water quality in tropical rivers in dry and rainy seasons: A model-based analysis (Under review).

Jerves-Cobo R, Lock K, Van Butsel J, Pauta G, Cisneros F, Nopens I, Goethals PLM. 2018. Integrated assessment of the chemical, microbial and ecological water quality in the Cuenca River basin (Ecuador). (In preparation)**.**

Conferences (oral presentations)

Montalvo, César A., **Jerves, Rubén**, Domínguez, Luis E. 2018. Determination of pollutant loads in the overflows of the combined sewage network of the City of Cuenca, Ecuador (Determinación de cargas de contaminación en aliviaderos de la red de alcantarillado combinado de la ciudad de Cuenca – Ecuador). XXXVI Inter-American Conference of Sanitary and Environmental Engineering (XXXVI Congreso Interamericano de Ingeniería Sanitaria y Ambiental). Guayaquil, Ecuador. October 28 – 31, 2018. http://www.aidisnet.org/html/esp/pub_mem-int.html

Jerves-Cobo, R., Iñiguez-Vela, X., Cordova-Vela, G., Diaz-Granda, C., Van-Echelpoel, W., Cisneros-Espinoza, F., Nopens, I. and Goethals, P. 2018. Model-based analysis of the potential of macroinvertebrates as indicators for microbial pathogens in rivers. Netherlands Annual Ecology Meeting. Netherlands Ecological Research Network. Lunteren, The Netherlands. February 13-14, 2018.

Rubio, Ricardo; Mora, Diego; **Jerves, Ruben**; Arias-Hidalgo, Mijail. 2017. Estimation of losses of precipitation in an urban watershed of the City of Cuenca (Estimación de las pérdidas de la precipitación en una cuenca hidrográfica urbana de la Ciudad de Cuenca)**.** First International Conference on Integrated Water Quality Management (IWQM). Cuenca, Ecuador. October 02–03, 2017.

Jerves-Cobo, R., Iñiguez-Vela, X., Cordova-Vela, G., Diaz-Granda, C., Van-Echelpoel, W., Cisneros-Espinoza, F., Nopens, I. and Goethals, P. 2016. Macroinvertebrate based mathematical models for the prediction of microbial pathogens in rivers**.** 3rd International Conference on Big Data Analysis and Data Mining. London, UK. September 26-27, 2016.
Journal of Computer Engineering & Information Technology, 2016, 5:4 (Suppl). doi: 10.4172/2324-9307.C1.005. ISSN: 2324-9307 JCEIT.

Jerves-Cobo, R., Zhindon-Argoti, D., Iñiguez-Vela, X., Cordova-Vela, G., Diaz-Granda, C., Cisneros-Espinoza, F., Nopens, I. and Goethals, P. 2015. Determination of physical–chemical conditions to predict macroinvertebrate communities in Machangara River (Southern Andes, in Ecuador). 21st International Congress on Modelling and Simulation. Gold Coast, Queensland, Australia. November 29 to December 4, 2015. ISBN 978-0-9872143-4-4, 288.
http://www.mssanz.org.au/modsim2015/documents/MODSIM2015_Abstracts.pdf

Educational activities

Promoter of master students

Montalvo Cesar. (2016-2018). Determination of pollutant loads in overflows of the combined sewage network of the City of Cuenca. (Determinación de cargas de contaminación en aliviaderos de la red de alcantarillado combinado de la ciudad de Cuenca). Cuenca University, Ecuador.

Rubio Ricardo. (2016-2018). Estimation of rainfall losses in an urban basin in the City of Cuenca (Estimación de las pérdidas de la precipitación en una cuenca hidrográfica urbana de la Ciudad de Cuenca). Cuenca University, Ecuador.

Grants

2015-2018	VLIR Ecuador Biodiversity Network
2015-2017	VLIR-UOS IUC Programme - University of Cuenca
2015	FWO Grant for participation at a congress in Gold Coast, Australia
1999-2000	AECI - Spanish International Cooperation Agency - Master scholarship.

Miscellaneous

2015-2018	Peer reviewer in scientific journals: Water (3 papers), Ecological Indicators (8 papers), Stats (1 paper), Science of the Total Environment (1 paper), Water Research (1 paper).

Appendix A – Supporting information for chapter 2

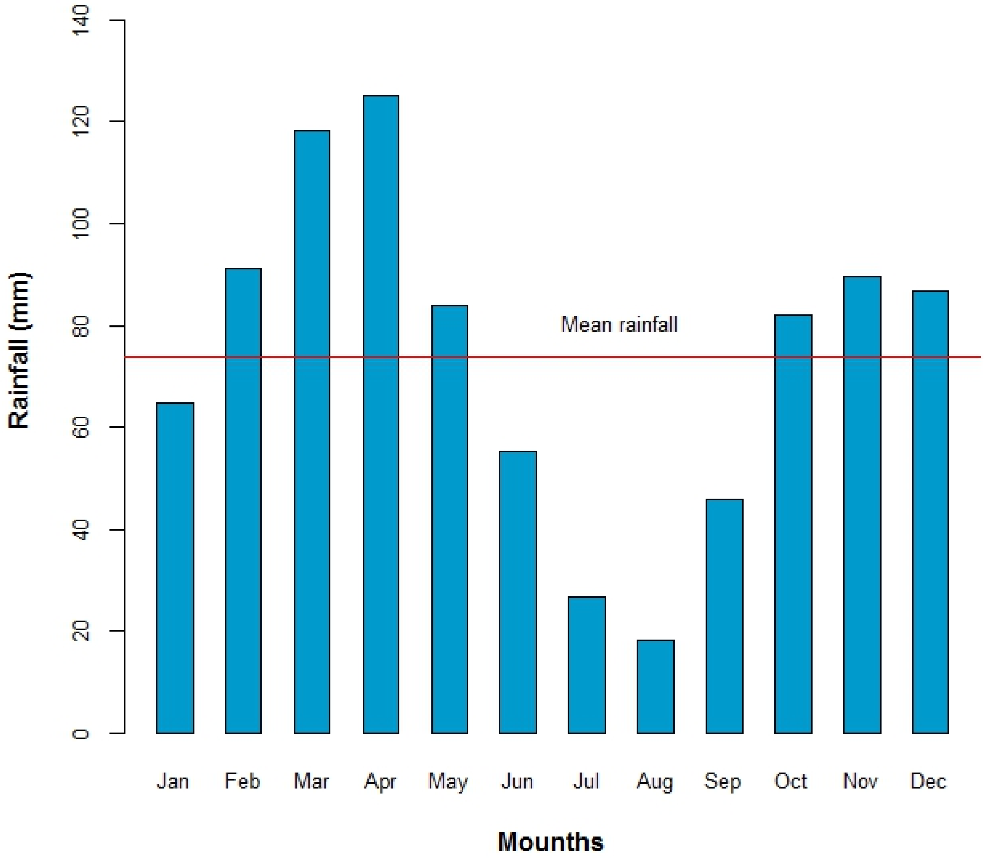


Fig. A1. Histogram of monthly average rainfall in the city of Cuenca (1990–2012)

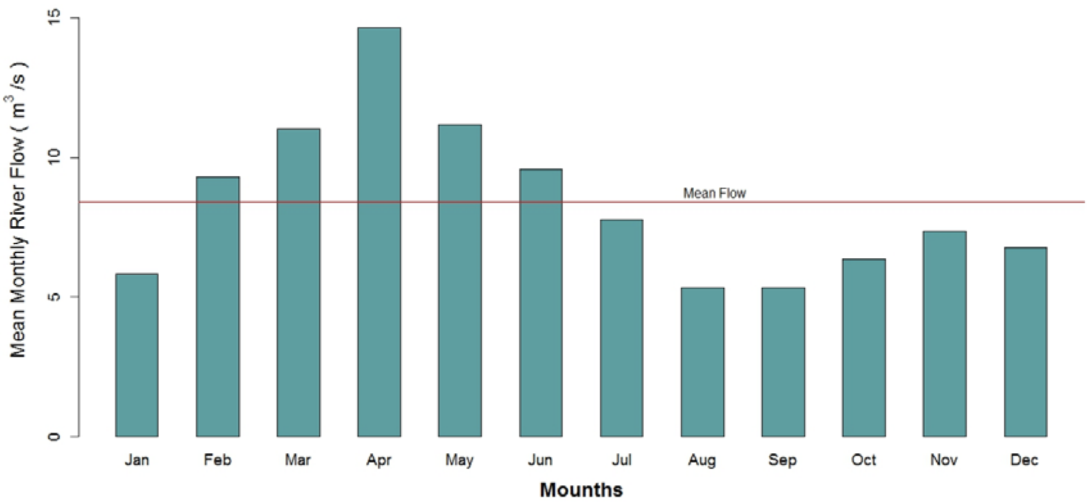


Fig. A2. Histogram Machangara River before discharging into Tomebamba River (1964–2010).

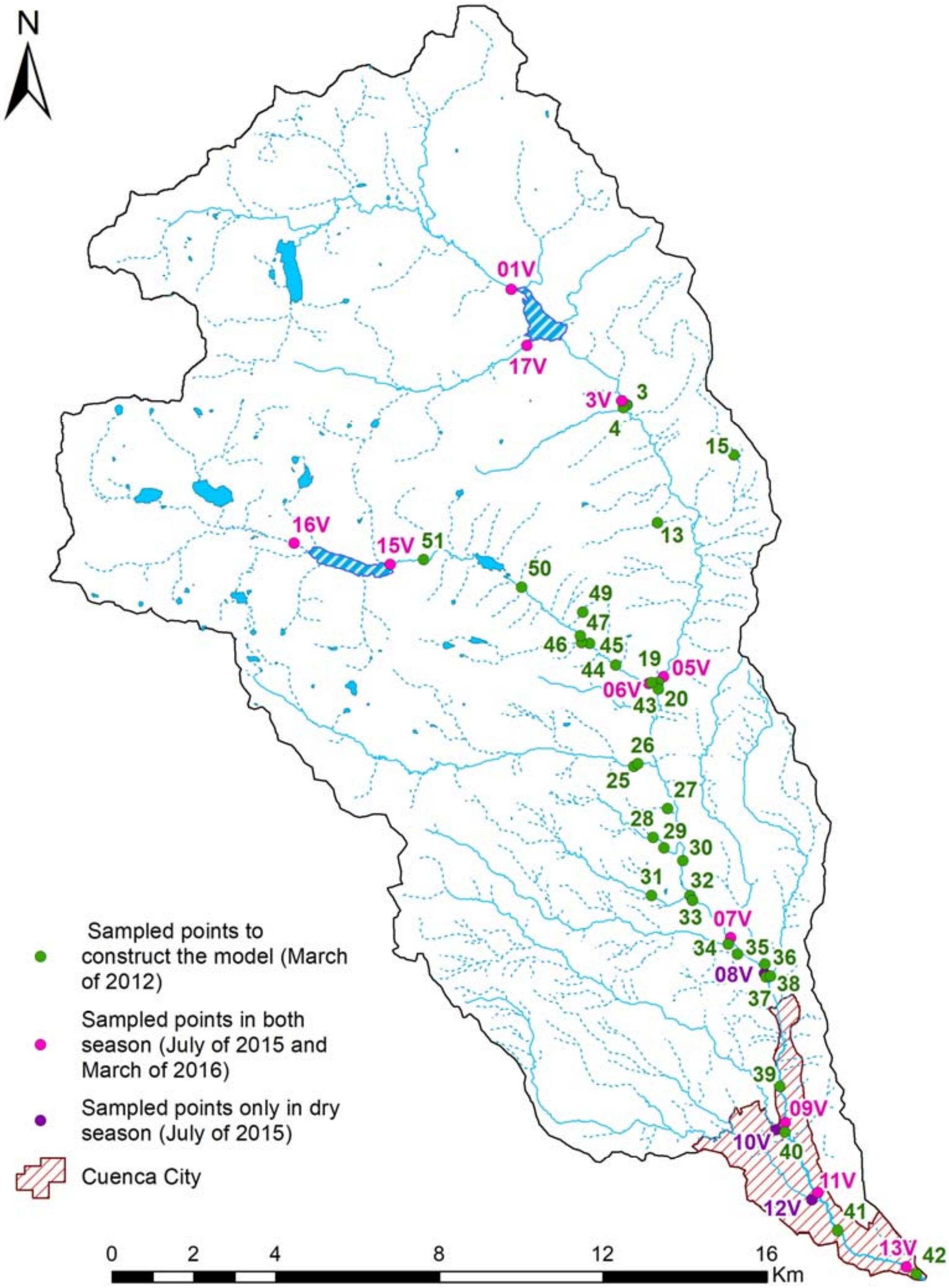
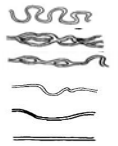
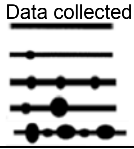


Fig. A3. Sampled sites location 2012, 2015 and 2016.

Table A1. Definition of categorical variables assessed in 43 sampling sites, modified from AusRivAS protocol (Parsons et al. 2002) and River Habitat Survey (Raven 1998).

Environmental variables	Number code (Categories)	Definition
Main land use (both banks for the stretch of 100m * 10m)	Forest	Land with high density of trees. Includes primary, secondary or tertiary forests.
	Arable	Land with agricultural crops (e.g. maize, vegetables)
	Residential	Land with residential houses
	Orchard	Land with fruit or nut-bearing trees
	Shrubs/grasses	Land with shrubs of grasses
Shading	No shading	No shading in the sampling site
	partly shaded, limited stretch <33%	Less than 33% of the sampling site is partly shaded
	partly shaded, longer stretch 33-90%	About 33 – 90 % of the sampling site is partly shaded
	partly shaded, whole stretch >90%	Greater than 90% of the sampling site is partly shaded
	completely shaded, limited stretch >33%	Less than 33% of the sampling site is completely shaded
	completely shaded, longer stretch 33-90%	About 33 – 90% of the sampling site is completely shaded
	completely shaded, whole stretch >90%	Greater than 90% of the sampling site is completely shaded
Main macrophytes	Absent	No macrophytes present
	Submerged macrophytes	Macrophytes rooted in the bottom sediment with the vegetative parts predominantly submerged
	Emerged macrophytes	Macrophytes rooted in the bottom sediment with vegetative parts emerging above the water surface
	Floating macrophytes	Macrophytes with roots, if present, hang free in the water and are not anchored to the bottom
Valley form	Canyon	
	V-shaped valley	
	Trough	
	Meander valley	
	U-shaped valley	
	Plain floodwidth	
Channel form	Meandering	
	Braided	
	Anabranching	
	Sinuate	
	Constrained (natural)	
	Constrained (artificial)	
Variation in width		Data collected at the reservoir 
Extent of erosion	Absent	Erosion is not visible
	Limited	Less than 30 % is eroded
	Abundant	More than 30% is eroded
Bank profile	Vertical	
	Steep	
	Gradually not trampled	
	Composite not trampled	
Variation of flow	Absent	No variation in flow
	At human constructions	Variation of flow at human construction
	Low	Variation of flow is less than 20%
	Moderate	Variation of flow is between 20 – 50%
	High	Variation of flow is greater than 50%
Depth of sludge layer	Absent	There is no sludge layer present.
	<5 cm	When the depth of sludge layer is less than 5 cm
	5-20 cm	When the depth of sludge layer is in between 5-20 cm
	>20 cm	When depth of sludge layer is greater than 20 cm
Abundance of dead wood	Absent	No dead wood present
Twigs (diameter <3 cm)	Limited	When dead wood present are less than two pieces
Branch (diameter 3-30 cm)		
Logs (diameter >30 cm)	Abundant	When dead wood present are more than three pieces
Pool/Riffle class	Class 1	Pool-riffle pattern is absent due to structural changes: uniform pool-riffle pattern due to reinforced bank and bed structures.
	Class 2	Pool-riffle pattern is absent: uniform pool-riffle pattern.
	Class 3	Pool-riffle pattern is poorly developed: low variety in pools and riffles.
	Class 4	Pool-riffle pattern is moderately developed: variety in pools and riffles but locally.

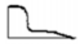
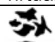
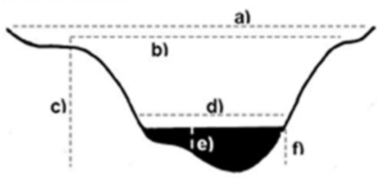
Environmental variables	Number code (Categories)	Definition
	Class 5	Pool-riffle pattern is well developed: high variety in pools and riffles.
	Class 6	Pool-riffle pattern is (nearly) pristine: extensive sequences of pools and riffles.
Bank shape	Concave	
	Convex	
	Stepped	
	Wide lower bench	
	Undercut	
Bank slope	Vertical	80-90° bank sloping
	Steep	60-80° bank sloping
	Moderate	30-60° bank sloping
	Low	10-30° bank sloping
	Flat	Less than 10° bank sloping
Bed compaction	Invisible	Bed not visible
	Tightly packed	Array of sediment sizes overlapping, tightly packed and very hard to dislodge
	Packed	Array of sediment sizes overlapping, tightly packed but can be dislodged moderately
	Moderate compaction	Array of sediment sizes, little overlapping, some packing but can be dislodged moderately
	Low compaction (1)	Limited range of sediment sizes, little overlapping, some packing and structure but can be dislodged very easily.
	Low compaction (2)	Loose array of fine sediments, no overlapping, no packing, and no structure and can be dislodge very easily.
Sediment matrix	Bedrock	Composed of bedrocks
	Open framework	0-5% fine sediment, high availability of interstitial space
	Matrix filled contact	5-32% fine sediments, moderate availability of interstitial space
	Framework dilated	32-60% fine sediments, low availability of interstitial space
	Matrix dominated	Greater than 60% fine sediments, interstitial space virtually absent
Sediment angularity	Very angular	
	Angular	
	Sub-angular	
	Rounded	
	Well rounded	
	Cobble, pebbles and gravel fractions not present	
Dominant mineral substrate (bed of the river)		Bedrock Boulder Cobble Pebble Gravel Sand Fines
Cross section measurement of the water course	Floodplain width Flood prone width Entrenchment depth Average stream width Average water depth Maximum water depth	
Presence of macroalgae	Absent	No macroalgae (mostly filamentous) present
	Present	Macroalgae is present
Presence of water bufallos	Absent	No domesticated water buffalos present
	Present	Domesticated water buffalos is present
Presence of sand/gravel quarrying activities	Absent	Sand/gravel quarrying are absent
	Present	Sand/gravel quarrying are present
Presence of water hyacinth	Absent	Water hyacinths are present
	Present	Water hyacinths are absent

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

			Tomebamba River (from upstream to downstream)																														Sampling sites with the same taxa
			Dry Season											Rainy Season																			
Site -> / Taxa	BMWP-Col Alvarez 2006	ABI Encalda 2011	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	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	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	4	2	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	p	3
Dytiscidae	-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Hydrophilidae	3	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Lymnaeidae	8	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	1
Physidae	3	3	-	-	p	-	-	-	p	-	p	p	p	-	-	p	p	p	-	p	p	p	p	p	p	p	p	p	p	p	-	p	20
Psychodidae	2	3	-	-	p	p	p	p	-	-	p	p	p	-	p	-	p	p	-	p	-	p	p	-	p	p	-	p	p	p	p	-	19
Sphaeriidae	8	3	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	2
Acari	-	4	p	-	p	-	-	-	-	-	-	-	-	-	p	p	p	-	-	p	p	-	p	-	-	-	p	p	-	-	-	-	10
Baetidae	7	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	5	4	p	-	p	-	-	-	-	-	p	-	-	-	-	p	-	-	-	p	p	p	-	-	-	-	-	p	-	-	-	-	8
Dolichopodidae	4	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Empididae	4	4	-	-	p	-	-	p	-	-	-	-	-	p	-	-	p	p	p	-	-	p	p	-	-	-	p	-	-	-	-	-	9
Pyralidae	7	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Tabanidae	5	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Elmidae	6	5	p	-	p	-	-	-	-	-	-	-	-	p	-	p	p	p	p	p	p	p	p	p	p	p	p	p	-	p	-	p	18
Hydropsychidae	7	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Limoniidae	-	5	p	p	p	-	-	-	p	-	-	-	-	p	p	p	p	p	p	p	p	p	p	-	p	-	-	-	-	-	-	p	16
Psephenidae	10	5	p	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	3
Scirtidae	4	5	-	-	-	-	-	-	-	-	p	-	-	p	-	p	p	p	p	p	p	p	-	-	-	p	-	-	-	-	-	-	10
Simuliidae	7	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	3	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	p	-	p	-	-	-	p	-	-	-	-	-	-	-	4
Aeshnidae	6	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Coenagrionidae	7	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Hyallelidae	7	6	p	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	p	p	p	-	p	p	-	p	-	p	p	-	p	11
Hydroptilidae	8	6	-	p	-	-	-	-	-	-	-	-	-	p	-	p	p	-	-	p	p	p	-	-	p	-	p	p	p	-	-	-	11
Limnephilidae	8	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	1
Glossosomatidae	7	7	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Leptohyphidae	7	7	p	-	p	-	-	-	-	-	p	-	-	p	p	p	-	-	-	p	p	-	p	-	p	-	p	-	-	-	-	-	11
Hydrobioscidae	9	8	p	-	-	-	-	-	-	-	-	-	-	p	-	p	p	-	-	p	p	p	p	p	p	-	p	p	p	p	-	-	14
Leptoceridae	8	8	-	-	-	-	-	-	-	-	-	-	-	-	-	p	p	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	3
Polycentropodidae	9	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Blepharoceridae	10	10	p	-	-	-	-	-	-	-	p	-	-	-	p	-	-	-	-	p	-	p	-	p	p	p	-	p	-	-	-	-	9
Calamoceratidae	8	10	p	-	-	-	-	-	-	-	-	-	-	-	p	p	p	-	-	p	p	p	-	p	p	-	p	-	-	-	-	-	10
Gripopterygidae	10	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	1
Helicopsychidae	8	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Leptophlebiidae	9	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Perlidae	10	10	-	-	-	-	-	-	-	-	-	-	-	p	-	p	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	3
Dugesiidae	6	-	p	-	p	-	-	-	p	-	-	-	-	-	-	p	-	p	-	p	-	p	p	-	-	p	p	p	p	-	-	p	13
Glossiphoniidae	5	-	-	-	-	-	-	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.

	Sensitivity Score		Machangara River (from upstream to downstream)								Yanuncay River (from upstream to downstream)												Tarqui River (from upstream to downstream)												Sampling sites with the same taxa in the three rivers
			Dry Season					Rainy Season			Dry Season				Rainy Season								Dry Season							Rainy Season					
Taxa \ Site ->	BMWP-Col Alvarez 2006	ABI Encalda 2011	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	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	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	4	2	-	p	p	p	-	-	-	p	-	-	-	-	-	p	-	-	-	p	-	p	-	-	-	-	p	-	-	-	-	-	-	-	8
Dytiscidae	-	3	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	2
Hydrophilidae	3	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	p	-	-	2
Lymnaeidae	8	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	2
Physidae	3	3	-	p	p	p	-	-	p	-	-	-	p	-	p	p	p	p	p	p	p	p	p	p	-	p	p	-	-	p	p	-	p	p	21
Psychodidae	2	3	-	p	-	p	p	p	-	p	-	-	-	p	-	-	-	p	p	p	p	p	-	p	-	-	-	p	-	-	-	-	p	p	15
Sphaeriidae	8	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	p	-	-	p	p	p	-	-	-	-	-	p	p	-	p	p	9
Acari	-	4	-	-	-	-	-	-	p	-	-	p	p	-	p	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	5
Baetidae	7	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	5	4	-	-	p	-	-	-	-	p	p	p	p	p	p	-	-	-	p	p	-	-	-	-	p	-	-	p	-	-	-	p	-	-	12
Dolichopodidae	4	4	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Empididae	4	4	-	p	p	-	-	-	-	-	-	-	p	-	p	-	-	-	-	p	-	-	-	p	-	-	-	-	p	p	p	-	-	-	9
Pyralidae	7	4	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Tabanidae	5	4	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Elmidae	6	5	p	p	p	-	-	-	p	p	p	p	p	p	p	p	-	-	-	-	-	-	-	-	p	-	-	p	-	-	p	p	p	-	16
Hydropsychidae	7	5	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Limoniidae	-	5	-	p	p	-	p	-	p	p	-	-	p	-	p	p	-	-	-	-	-	p	-	p	p	-	-	p	-	-	p	p	-	p	15
Psephenidae	10	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Scirtidae	4	5	-	p	-	-	-	-	-	-	-	p	-	-	p	p	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	p	-	-	6
Simuliidae	7	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	3	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	p	-	-	-	-	2
Aeshnidae	6	6	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	p	-	-	-	-	3
Coenagrionidae	7	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	1
Hyallelidae	7	6	-	p	-	p	-	-	-	-	p	-	p	p	-	p	p	p	-	p	p	p	p	p	-	p	-	-	p	p	p	p	p	p	20
Hydroptilidae	8	6	p	p	-	-	-	-	p	p	-	-	-	-	-	-	-	-	-	p	-	p	-	-	-	p	-	-	-	-	p	p	-	-	9
Limnephilidae	8	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Glossosomatidae	7	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Leptohyphidae	7	7	-	-	-	-	-	-	-	-	p	p	p	-	p	p	-	p	-	-	-	p	-	p	p	-	-	-	-	-	-	p	-	-	10
Hydrobioscidae	9	8	p	p	p	-	p	-	-	p	-	p	p	-	-	-	-	-	-	-	-	-	-	p	p	-	-	-	-	-	p	p	-	-	11
Leptoceridae	8	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
Polycentropodidae	9	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	p	-	-	2
Blepharoceridae	10	10	p	-	p	-	-	-	-	-	-	-	p	p	p	p	-	p	-	p	-	-	-	-	-	-	-	-	-	-	-	p	-	-	9
Calamoceratidae	8	10	-	-	-	-	-	-	-	-	-	-	-	-	p	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	3
Gripopterygidae	10	10	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Helicopsychidae	8	10	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Leptophlebiidae	9	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	2
Perlidae	10	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	1
Dugesiidae	6	-	-	-	-	p	-	-	p	-	-	-	-	-	-	-	-	p	p	-	p	p	p	p	-	p	p	p	p	p	-	p	p	p	16
Glossiphoniidae	5	-	-	-	-	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
Stenopsychidae	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
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 Deficie; 5 Bad

Table A4. Taxa present in sampling sites in the Machangara River during February and March of 2012.

Site -> / Taxa	BMWP-Col Sensitivity Score (1)	Sampling sites																																	Sampling sites with the same taxa
		03	04	13	15	19	20	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	49	50	51	
Water quality ->		2	3	2	3	4	3	3	2	3	2	2	3	3	4	4	2	3	3	3	4	3	3	4	5	5	2	3	3	2	3	2	4	5	
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	-	27
Dytiscidae	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	1
Physidae	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	p	-	-	-	-	-	-	-	-	-	-	-	p	3
Glossiphoniidae	3	-	-	-	-	-	-	-	p	-	-	-	-	p	-	-	-	p	p	-	p	-	-	-	p	-	-	-	-	-	-	-	-	-	6
Empididae	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Psychodidae	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Hydrophilidae	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	1
Planorbidae	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	p	-	-	-	2
Dixidae	4	-	-	-	-	-	-	-	-	-	p	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
Limoniidae	4	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	p	-	-	-	-	-	-	-	3
Tabanidae	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	p	-	3
Limnaeidae	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	p	-	p	-	p	-	4
Ceratopogonidae	4	p	-	-	-	p	p	-	p	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	6
Planariidae	4	-	-	-	-	p	-	-	-	-	-	-	-	-	p	-	-	-	-	p	-	-	-	p	-	-	-	-	-	p	p	p	p	-	8
Staphylinidae	5	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
Hydropsychidae	5	p	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	3
Muscidae	5	-	-	-	p	-	p	p	p	p	-	p	-	-	-	-	p	p	-	p	-	-	-	-	-	-	-	-	-	p	-	-	-	-	10
Tipulidae	5	-	p	p	p	p	-	-	-	p	p	-	-	p	-	-	p	p	-	-	p	-	p	-	-	-	p	-	p	-	p	-	-	-	14
Baetidae	5	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	-	-	27
Scirtidae	6	-	-	-	p	p	p	-	-	p	p	p	-	-	p	-	p	p	p	p	-	p	-	-	-	-	-	-	-	-	-	-	-	-	12
Elmidae	6	p	p	-	p	-	-	p	p	-	p	p	p	p	p	-	p	p	-	p	-	-	-	-	-	-	p	-	-	p	p	p	-	-	17
Simuliidae	6	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	-	-	-	27
Aeshnidae	7	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	2
Tricorythidae	7	p	p	-	-	-	p	p	p	p	p	-	-	-	-	-	p	-	-	-	p	p	-	-	-	-	p	-	p	p	-	-	-	-	13
Hyalellidae	7	p	p	-	p	p	p	-	p	p	p	p	-	p	-	-	p	-	p	p	-	p	-	-	-	-	p	p	p	p	p	p	p	p	22
Leptoceridae	8	-	-	p	-	-	-	-	-	-	-	-	p	-	-	p	-	-	-	-	-	-	-	p	-	-	-	-	p	-	-	p	-	-	6
Leptophlebiidae	8	p	-	p	-	-	-	-	p	p	p	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	p	-	p	-	-	9
Hydrobiosidae	8	p	p	p	p	-	-	p	p	-	-	p	p	-	-	-	-	p	-	-	-	p	p	p	-	-	-	p	p	p	p	p	-	-	17
Helicopsychidae	9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	1
Ptilodactylidae	9	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	2
Odontoceridae	9	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	2
Hydracarina	10	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
Blepharoceridae	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	p	-	-	-	-	2
Gripopterygidae	10	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	2
Calamoceratidae	10	p	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	p	-	-	4
Perlidae	10	-	p	p	-	-	-	p	p	-	p	p	p	p	-	-	p	-	-	-	-	-	-	-	-	-	p	-	-	p	-	p	-	-	12
Total taxa per point →		12	10	11	10	8	10	9	14	10	13	10	8	7	6	4	12	12	7	11	8	11	7	7	4	3	12	7	11	14	11	14	7	3	

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

Taxa considered in the construction of the models→ (blue shading)

p = Taxa present

(1) BMWP-Col Cajas

Table A5. Abundance of the taxa present in sampling sites in the Machangara River during February and March of 2012.

		Sampled Points in the Machangara River																																
Site / Taxa	Sensitivity Score ↓	3	4	13	15	19	20	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	49	50	51	Points with the Same Taxa
Biological Water Quality BMWP-Col →		69	57	71	59	35	48	58	82	55	74	63	46	40	29	19	64	55	39	49	33	59	38	34	11	8	67	37	57	82	56	93	29	11	
Land use category →		5	5	5	3	5	4	3	5	6	3	6	6	3	6	3	6	2	2	2	2	2	1	1	1	1	5	5	5	3	6	5	3	5	
Tubificidae	1	38	1	1		1	1	1	1	1	1		1			1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	28
Chironomidae	2	38	4	1	56	9	4		6		2	4	4		2		1	15	2	40	2	6		9	43	15	8	2	3	2	23	10	4		27
Dytiscidae	3																													1					1
Physidae	3																			14		1												1	3
Glossiphoniidae	3								1					1				1	1		3				1										6
Empididae	4																	1																	1
Psychodidae	4															1																			1
Hydrophilidae	4																						1												1
Planorbidae	4																												1		2				2
Dixidae	4										1			1																					2
Limoniidae	4				3																	5					3								3
Tabanidae	4																1			1													1		3
Limnaeidae	4																				1								1		2		2		4
Ceratopogonidae	4	1				1	1		2									1									1								6
Planariidae	4					6									1					14				2						1	6	1	1		8
Staphylinidae	5									2																									1
Hydropsychidae	5	1					1																									1			3
Muscidae	5				1		2	1	5	5		2					3	2		3										1					10
Tipulidae	5		1	6	4	5				2	1			3			1	1			1		3				1		1		1				14
Baetidae	5	3	13	37			2	43	16	2	38	3	2	1	1		3	53	7	189	45	108	151	31	17	19	1	3	4	30		140			27
Scirtidae	6				132	1	1			2	3	1			1		1	1	4	4		2													12
Elminthidae	6	2	2		3			16	2		3	1	1	5	1		2	4		1							1			4	1	1			17
Simuliidae	6	18	3	3	13	1	21	2	31	26	19	45	8		18	5	14	51	22	10	1	14	2	1			18	1	5	3	15				27
Aeshnidae	7										1																						1		2
Tricorythidae	7	25	7				6	2	3	2	8						1				1	14					9		1	2					13
Hyalellidae	7	88	4		6	30	7		10	1	16	3		2			3		2	1		1					6	2	2	2	116	61	38	5	22
Leptoceridae	8			1									1			1								1					1			2			6
Leptophlebiidae	8	10		4					1	2	6	5																1		2		56			9
Hydrobiosidae	8	2	3	6	4			2	1			9	1					1				14	2	2				1	1	1	1	6			17
Helicopsychidae	9																															1			1
Ptilodactylidae	9			1																			1												2
Odontoceridae	9			1																											1				2
Hydracarina	10				1														1																2
Blepharoceridae	10																										1			1					2
Gripopterygidae	10								1																							1			2
Calamoceratidae	10	2						1														1										1			4
Perlidae	10		1	7				6	1		4	1	1	1			1										1			1		9			12
Taxa richness →		12	10	11	10	8	10	9	14	10	13	10	8	7	6	4	12	12	7	11	8	11	7	7	4	3	12	7	11	14	11	14	7	3	

Land use category →	1	Urban area	2	Suburban area, pastures and crops	3	Pastures	4	Bare soil	5	Native vegetation	6	Forest vegetation
Biological water quality classification →		Very good		Good		Moderate		Deficient		Bad		
Taxa analyzed to construct the models →		Included				Not included	p = Taxa present					

Table A6. Taxa present in sampling sites in the Machangara River in the non-urban area, during June of 2015 and March of 2016 (dry and rainy seasons).

	Sensitivity Score		**Dry Season**									**Rainy Season**							
Site -> Taxa	BMWP-Col Alvarez 2006	ABI Encalda 2011	MC01	MC17	MC03	MC05	MC16	MC15	MC06	MC07	MC08	MC01R	MC17R	MC03R	MC05R	MC16R	MC15R	MC06R	MC07R
Water quality ->			3	2	3	3	2	5	3	3	3	3	3	3	3	3	3	3	4
Tubificidae	1	1	-	-	p	-	p	p	-	-	p	p	p	p	p	p	-	p	p
Chironomidae	2	2	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	p	-
Hydrophilidae	3	3	-	-	p	-	-	-	-	p	-	-	-	-	-	-	-	-	-
Lymnaeidae	8	3	-	-	-	-	-	p	-	-	-	-	-	p	-	-	p	p	-
Physidae	3	3	-	-	p	-	-	-	-	-	p	-	-	-	-	-	-	-	p
Planorbidae	8	3	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-
Psychodidae	2	3	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-
Sphaeriidae	8	3	-	-	-	-	-	-	p	-	-	p	-	p	-	-	-	-	-
Acari	-	4	-	p	-	-	-	-	p	p	-	p	-	-	-	-	-	p	-
Baetidae	7	4	p	p	-	p	p	-	p	p	p	p	-	p	p	-	p	p	p
Ceratopogonidae	5	4	p	-	p	-	p	-	-	-	-	p	-	p	-	-	p	-	-
Dolichopodidae	4	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Empididae	4	4	-	-	-	-	p	-	-	-	p	-	-	p	p	-	-	p	-
Tabanidae	5	4	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Elmidae	6	5	p	p	p	-	p	-	p	p	p	p	p	p	p	p	-	-	p
Hydropsychidae	7	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Limoniidae	-	5	-	-	-	p	p	-	p	p	p	p	-	-	p	p	p	p	p
Scirtidae	4	5	p	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Simuliidae	7	5	-	p	p	p	p	-	p	p	p	p	p	p	p	p	-	p	p
Tipulidae	3	5	-	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-
Hyallelidae	7	6	p	p	p	p	p	p	p	-	-	p	-	p	p	p	p	p	-
Hydroptilidae	8	6	-	-	-	p	-	-	p	-	-	-	p	p	p	-	p	p	-
Limnephilidae	8	6	-	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-
Leptohyphidae	7	7	-	-	-	-	p	-	-	-	-	-	-	-	-	p	-	-	-
Hydrobioscidae	9	8	p	p	-	p	p	-	-	p	p	p	p	-	-	p	-	-	-
Leptoceridae	8	8	-	p	p	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Polycentropodidae	9	8	p	-	-	-	-	-	-	-	-	-	-	-	-	-	p	-	-
Blepharoceridae	10	10	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-
Calamoceratidae	8	10	-	p	p	-	-	-	-	-	-	-	-	p	-	-	-	-	-
Gripopterygidae	10	10	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-	-	-
Leptophlebiidae	9	10	p	p	-	-	p	-	-	-	-	p	p	-	-	p	-	-	-
Perlidae	10	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Dugesiidae	6	-	-	-	-	p	p	-	p	-	p	-	-	p	p	-	p	p	-
Glossiphoniidae	5	-	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-
Lumbricidae	-	-	p	p	p	-	p	-	p	p	p	p	p	p	p	p	p	p	-
Sericostomatidae	-	-	-	p	-	-	p	-	-	-	-	-	-	-	-	p	-	-	-
Stenopsychidae	-	-	-	-	-	-	-	-	p	-	-	-	-	-	-	-	-	-	-
Total taxa per point →			10	13	12	8	16	4	12	9	14	13	9	15	11	12	10	12	6
p = Taxa present																			
Biological water quality classification →					1	Very good		2	Good		3	Moderate		4	Deficie		5	Bad	

Appendix B – Supporting information for chapter 3

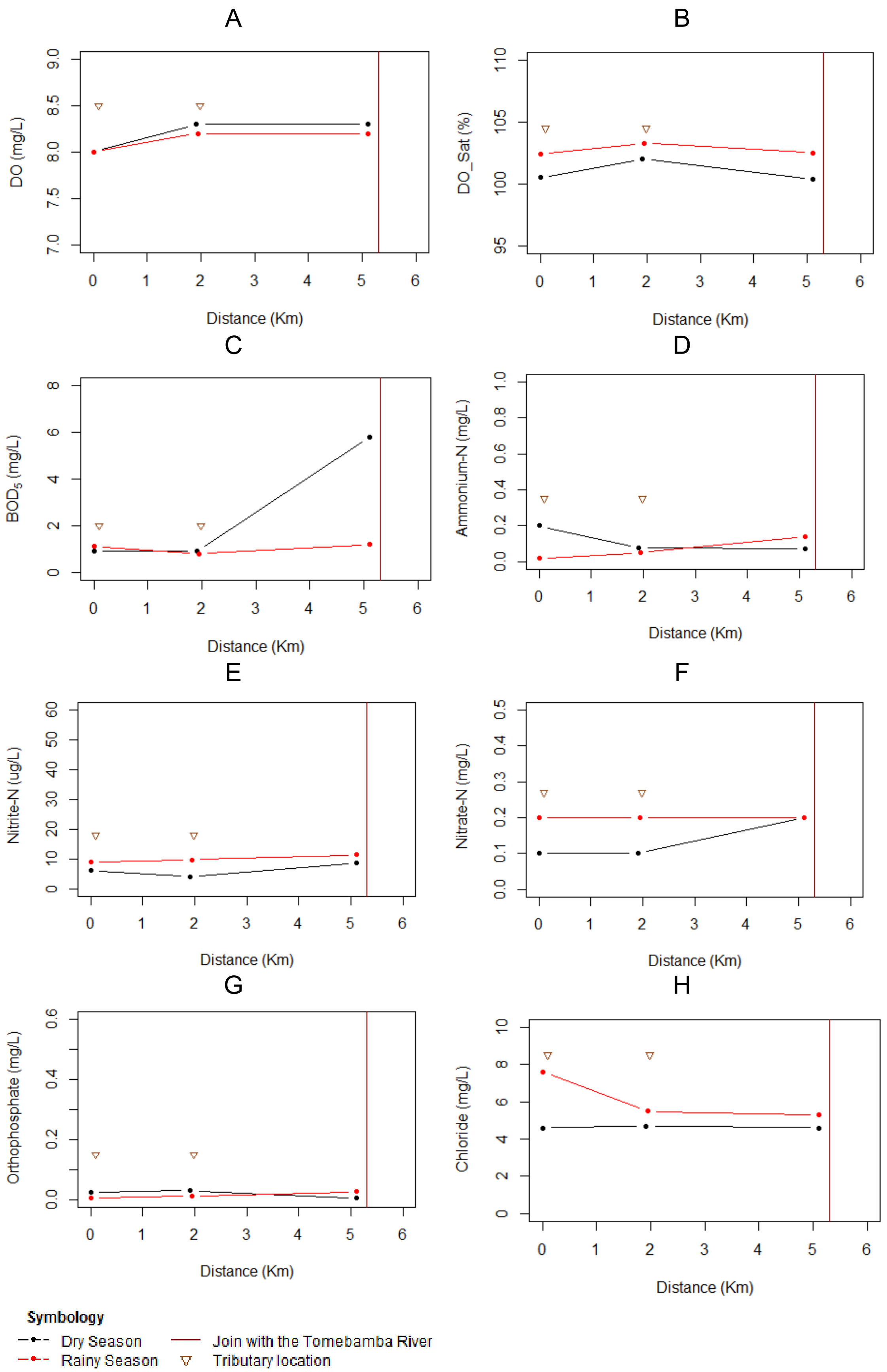


Fig. B1. Parameters variation along the Machangara River: (**A**) DO, (**B**) DO-Saturation, (**C**) BOD_5, (**D**) Ammonium, (**E**) Nitrite, (**F**) Nitrate, (**G**) Orthophosphate, (**H**) Chloride.

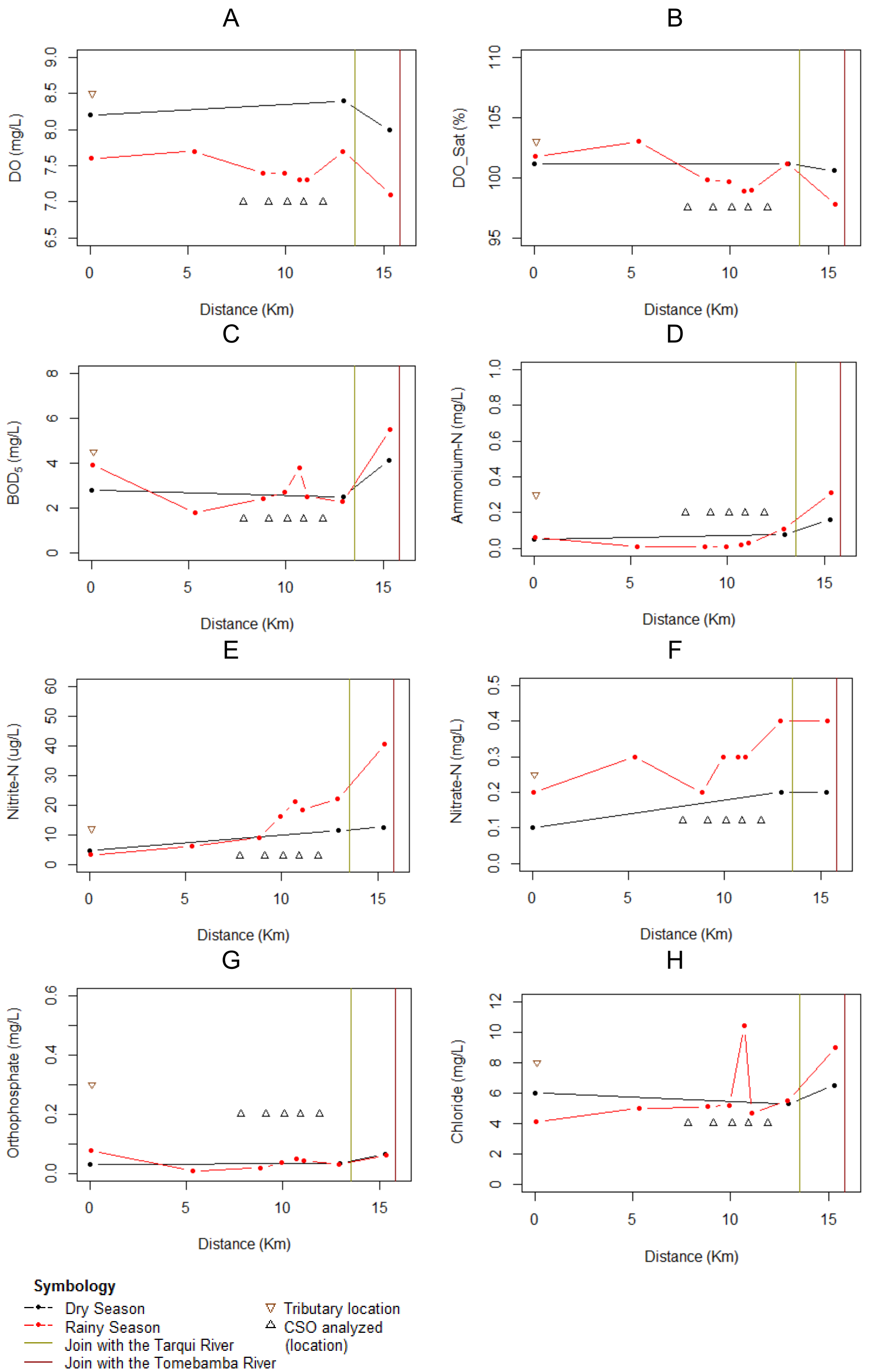


Fig. B2. Parameters variation along the Yanuncay River: (**A**) DO, (**B**) DO-Saturation, (**C**) BOD_5, (**D**) Ammonium, (**E**) Nitrite, (**F**) Nitrate, (**G**) Orthophosphate, (**H**) Chloride.

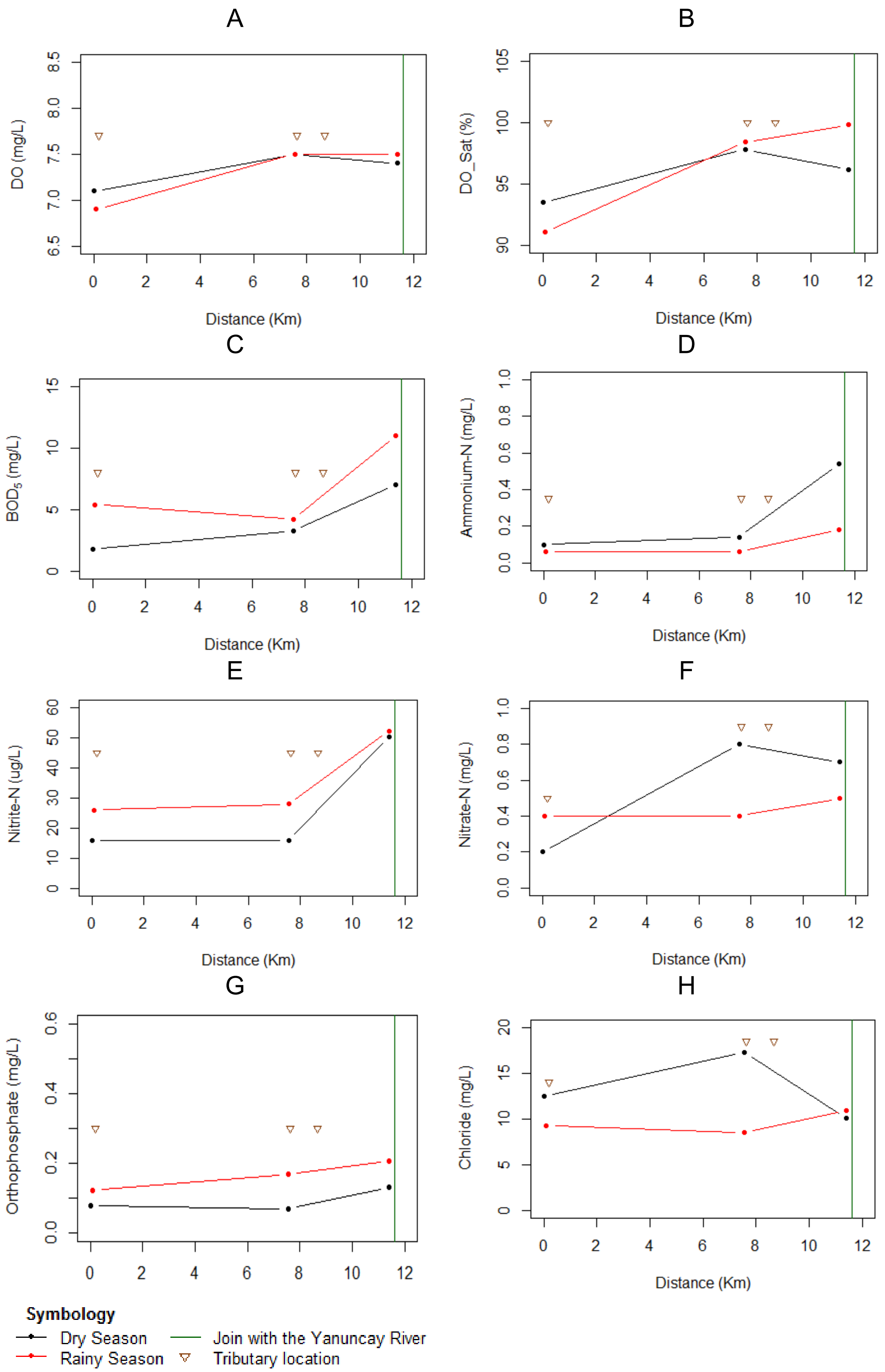


Fig. B3. Parameters variation along the Tarqui River: (**A**) DO, (**B**) DO-Saturation, (**C**) BOD_5, (**D**) Ammonium, (**E**) Nitrite, (**F**) Nitrate, (**G**) Orthophosphate, (**H**) Chloride.

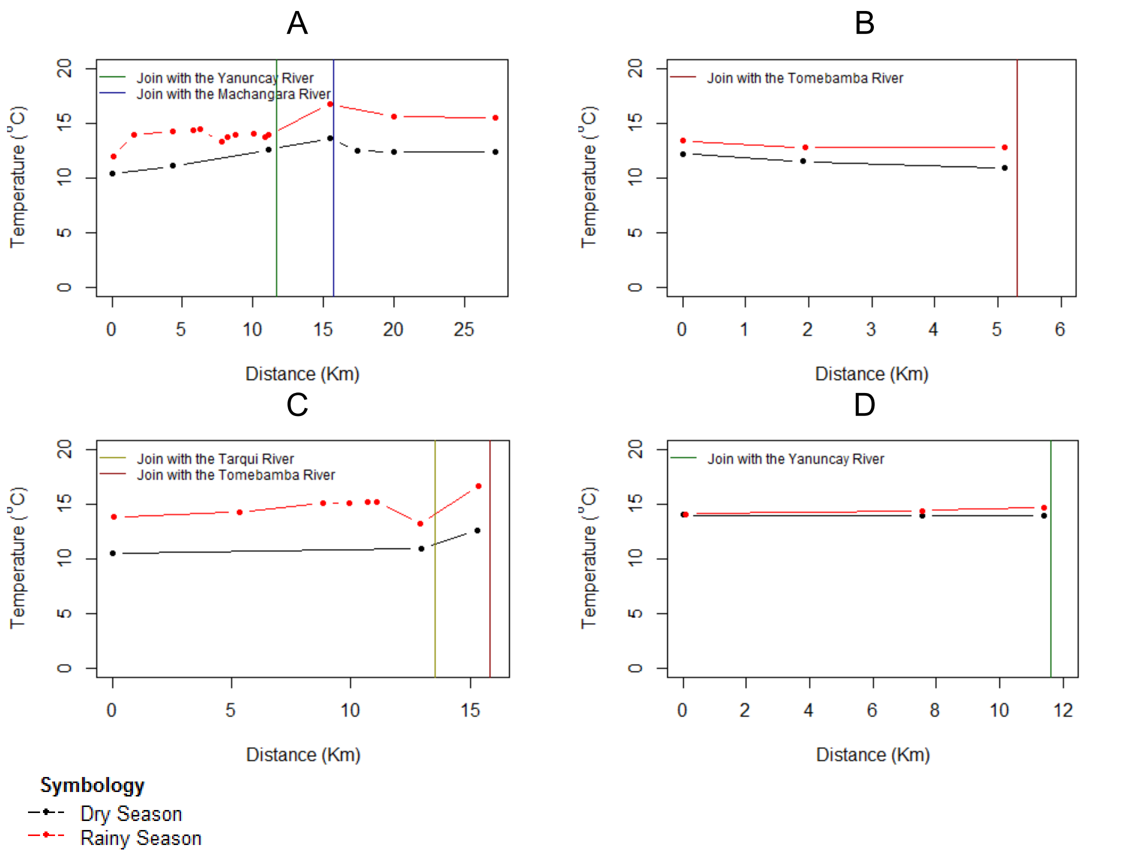


Fig. B4. Temperature variation in dry and rainy season along the rivers: (A) Tomebamba, (B) Machangara, (C) Yanuncay and (D) Tarqui.

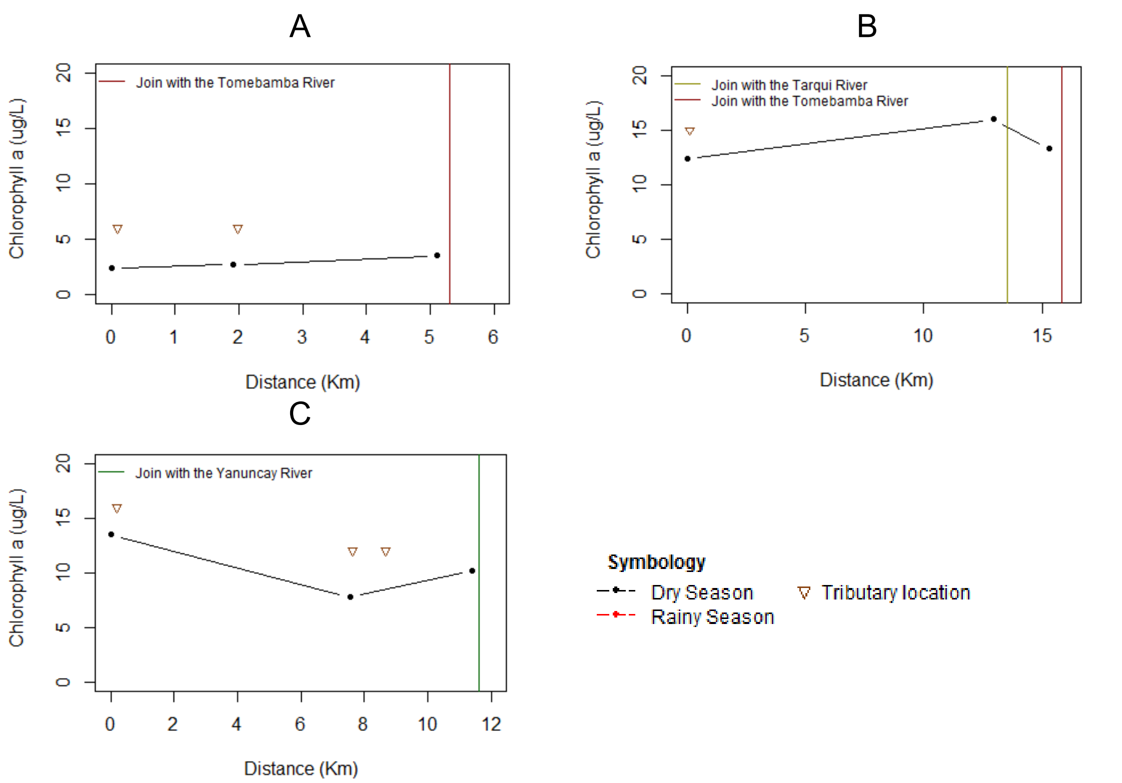


Fig. B5. Chlorophyll-a variation in dry and rainy season along the rivers: (**A**) Machangara, (**B**) Yanuncay and (**C**) Tarqui.

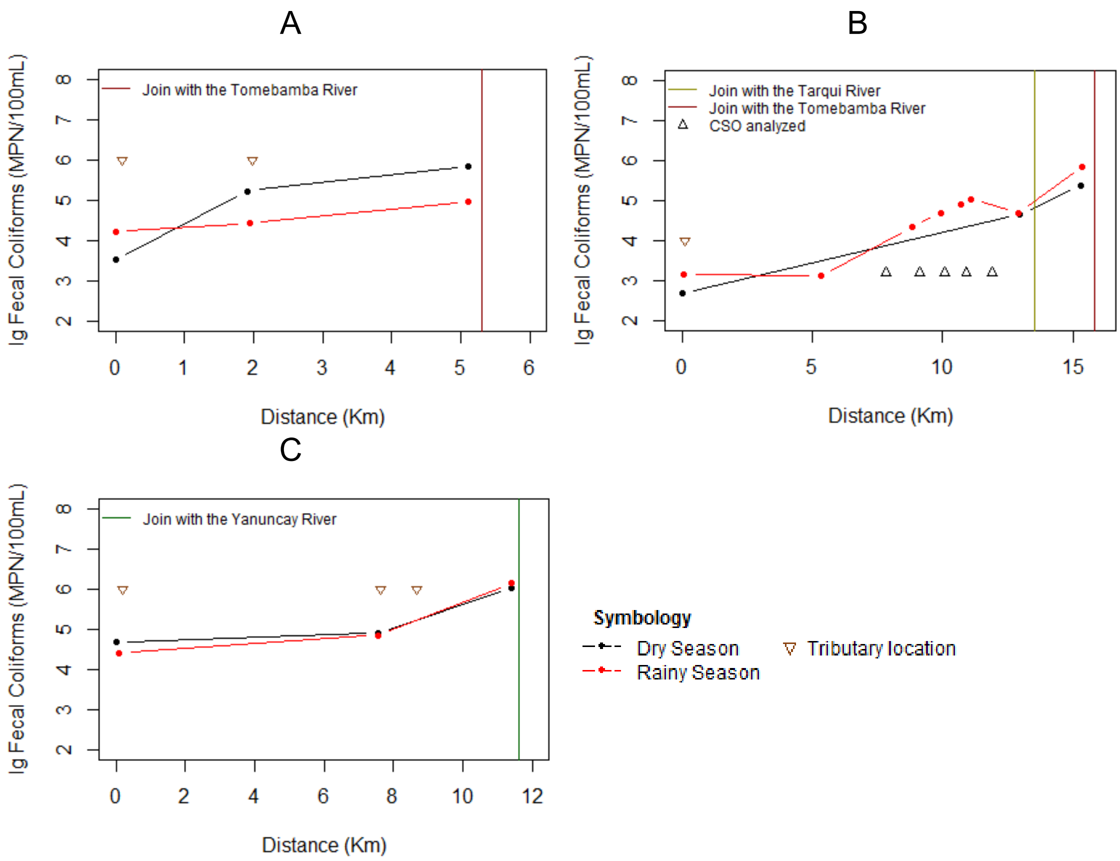


Fig. B6. Fecal coliforms variation in dry and rainy season along the rivers: (**A**) Machangara, (**B**) Yanuncay and (**C**) Tarqui.

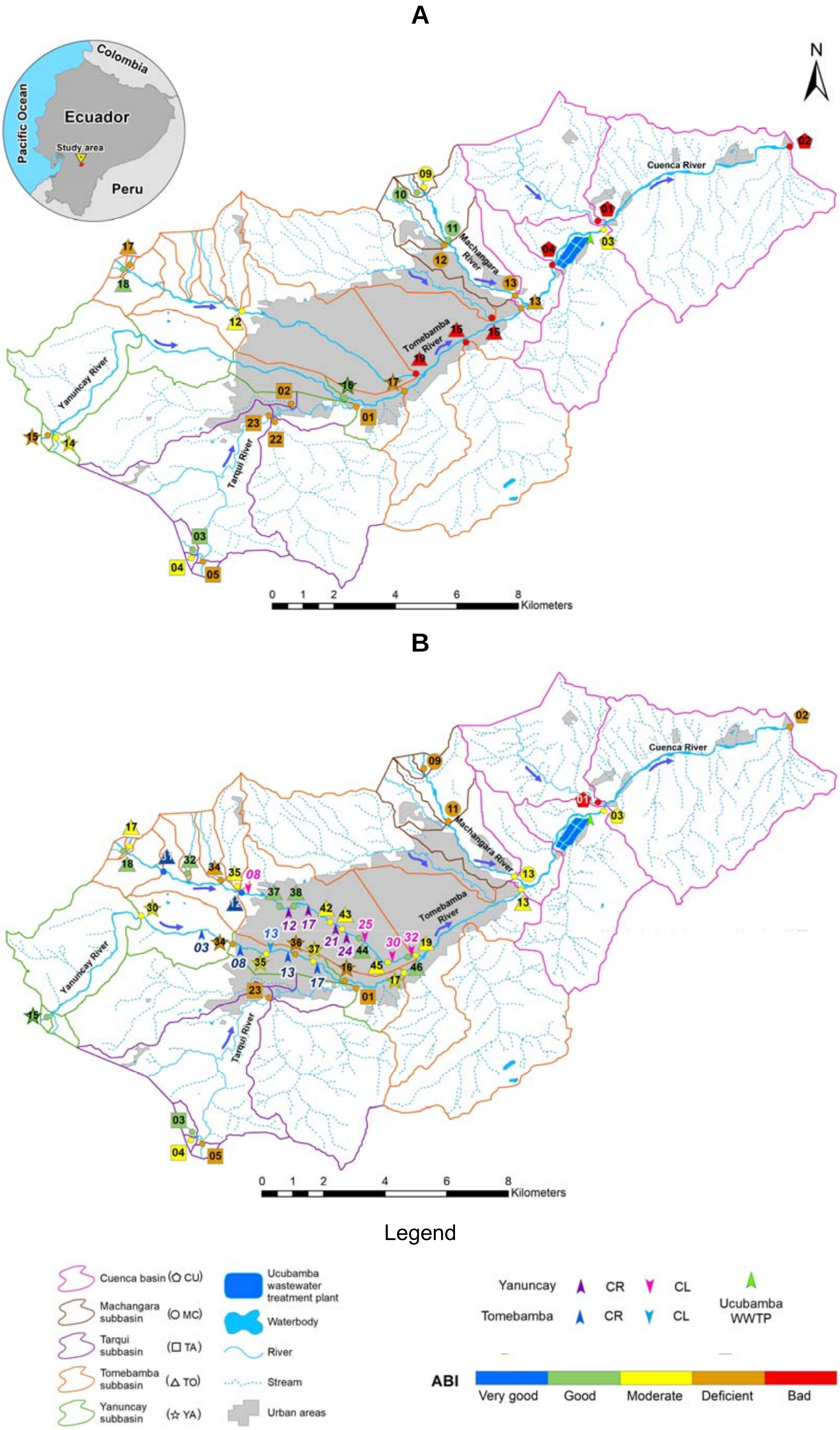


Fig. B7. Sampling sites location with their Andean Biotic Index (ABI) in both seasons: **(A)** dry season, (**B**) rainy season.

Table B1. Abundance analysis in the outfalls and the urban area: mean and standard deviation

	Baetidae		Ceratopogonidae		Chironomidae		Dugesiidae		Elmidae		Glossiphoniidae		Hyallelidae		Lumbricidae		Psychodidae		Physidae		Simuliidae		Tubificidae	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD
Outfalls: upstream dry season	89	138	4	2	10	7	17	22	2	1	1	0	6	9	10	14	2		4		6	9	21	40
Outfalls: downstream dry season	153	217	2	1	54	46	7	6	2	1	15	14	3	1	10	15	4	5	4	4	4	1	237	265
Outfalls: total dry season	121	177	2	2	32	39	11	13	2	1	9	12	5	7	10	13	4	4	4	4	5	7	139	220
Outfalls: upstream rainy season	66	79	11	21	31	28	10	18	8	15	4	3	97	231	6	5	4	3	105	212	23	42	209	266
Outfalls: downstream rainy season	74	58	2	1	62	124	8	13	5	5	29	50	9	13	14	18	7	9	49	64	24	33	72	61
Outfalls: total rainy season	70	68	7	16	47	91	9	15	7	12	19	40	43	145	11	14	6	7	76	152	24	37	140	201
Outfalls: upstream both seasons	74	99	9	18	24	25	12	17	7	14	3	3	70	194	7	8	4	3	96	204	19	37	150	236
Outfalls: downstream both seasons	101	132	2	1	60	104	8	11	4	4	25	43	8	12	13	17	6	8	38	59	21	31	130	176
Outfalls: total both seasons	87	116	6	13	42	78	10	14	6	10	16	35	35	128	11	14	5	6	63	140	20	34	140	204
Urban area: abundance dry season	69	133	4	6	95	234	6	10	2	1	6	9	14	34	7	10	10	22	53	53	4	6	123	246
Urban area: abundance rainy season	61	66	7	15	163	705	8	13	5	10	17	36	37	134	10	14	5	6	140	140	20	33	108	180
Urban area: abundance both seasons	65	101	5	11	133	547	7	12	4	8	13	29	30	112	9	12	7	14	121	121	15	28	114	210

Table B2. Abundance analysis in the outfalls and the urban area: min and max

	Baetidae		Ceratopogonidae		Chironomidae		Dugesiidae		Elmidae		Glossiphoniidae		Hyallelidae		Lumbricidae		Psychodidae		Physidae		Simuliidae		Tubificidae	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
Outfalls: upstream dry season	1	355	2	5	1	20	1	32	1	2	1	1	1	17	1	30	2	2	4	4	1	19	1	92
Outfalls: downstream dry	8	552	1	3	8	131	1	13	1	3	2	29	2	3	1	32	1	10	1	10	3	5	10	725
Outfalls: total dry season	1	552	1	5	1	131	1	32	1	3	1	29	1	17	1	32	1	10	1	10	1	19	1	725
Outfalls: upstream rainy season	1	289	1	49	5	98	1	49	1	48	2	9	1	621	1	14	1	8	2	696	1	138	1	796
Outfalls: downstream rainy	1	180	1	4	2	469	1	39	1	15	1	137	1	45	1	69	1	29	1	221	1	101	6	221
Outfalls: total rainy season	1	289	1	49	2	469	1	49	1	48	1	137	1	621	1	69	1	29	1	696	1	138	1	796
Outfalls: upstream both	1	355	1	49	1	98	1	49	1	48	1	9	1	621	1	30	1	8	2	696	1	138	1	796
Outfalls: downstream both	1	552	1	4	2	469	1	39	1	15	1	137	1	45	1	69	1	29	1	221	1	101	6	725
Outfalls: total both seasons	1	552	1	49	1	469	1	49	1	48	1	137	1	621	1	69	1	29	1	696	1	138	1	796
Urban area: abundance dry	1	552	1	20	1	1102	1	32	1	5	1	29	1	109	1	32	1	81	1	182	1	19	1	1007
Urban area: abundance rainy	1	289	1	49	1	4065	1	49	1	48	1	137	1	621	1	69	1	29	1	696	1	138	1	796
Urban area: abundance both	1	552	1	49	1	4065	1	49	1	48	1	137	1	621	1	69	1	81	1	696	1	138	1	1007

Appendix C – Supporting information for chapter 4

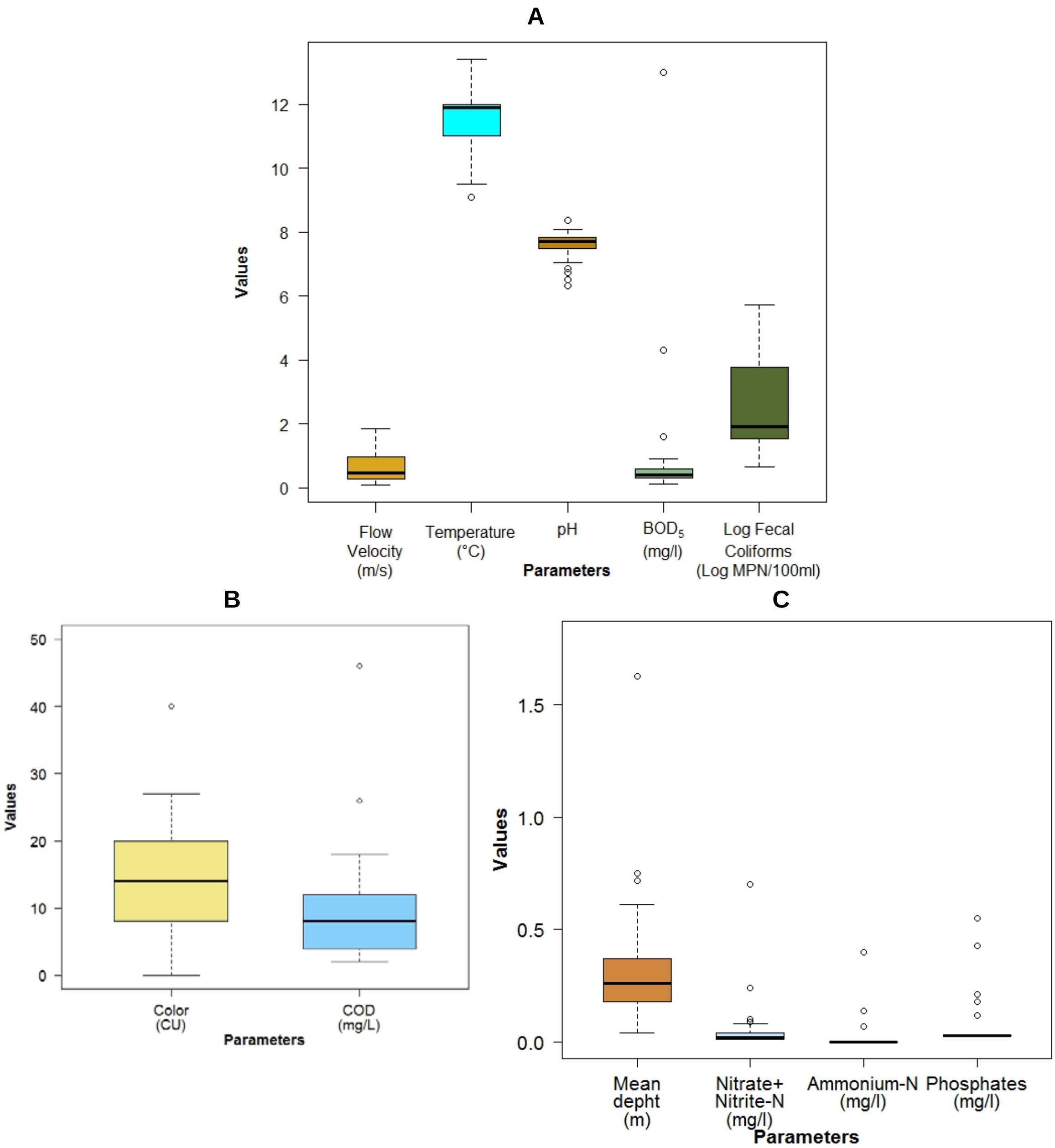


Fig. C1. Boxplots of the physicochemical parameters: (**A**) Flow Velocity, Water Temperature, pH, BOD_5, Log Fecal Coliforms; (**B**) True color and COD; (**C**) Mean depth, Nitrate + Nitrite, Ammonium and Phosphates. Dissolved Oxygen (DO) and Organic Nitrogen.

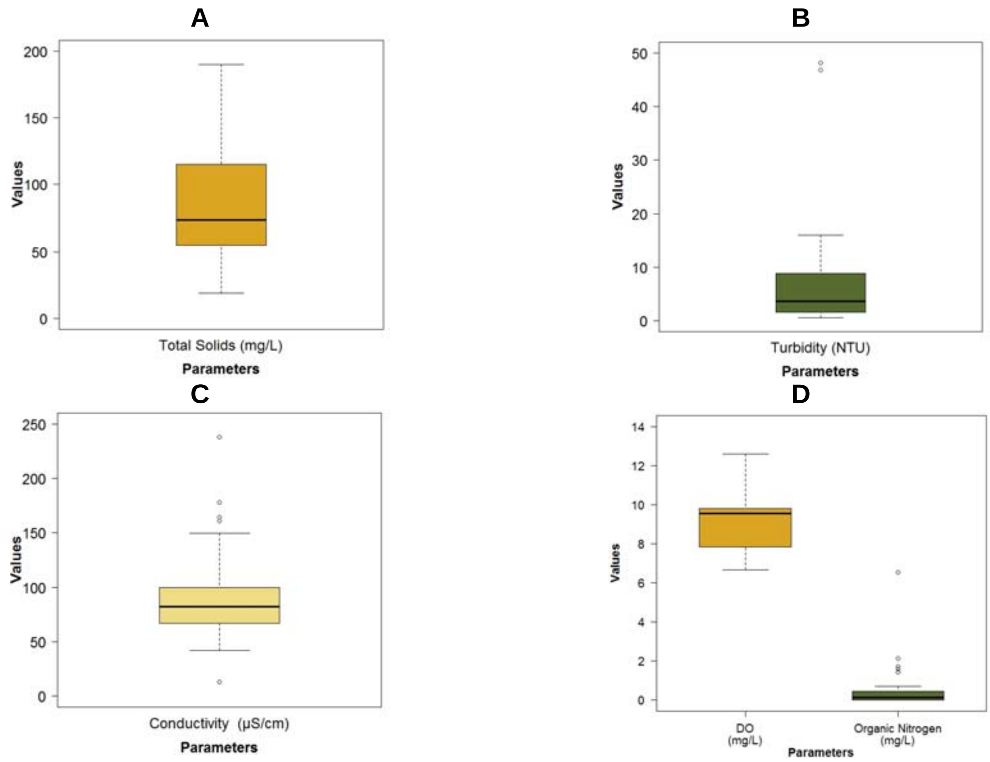


Fig. C2. Boxplots of the physicochemical parameters: (**A**) Total Solids; (**B**) Turbidity; (**C**) Conductivity; (**D**) Dissolved Oxygen (DO) and Organic Nitrogen.

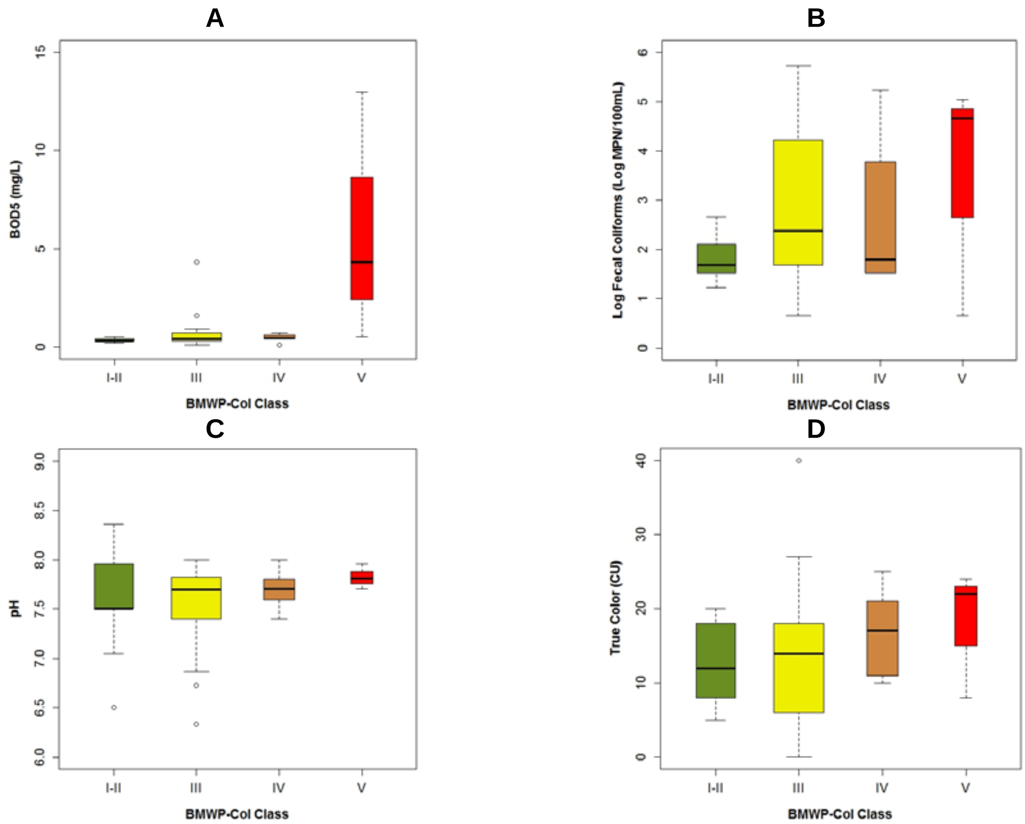


Fig. C3. Boxplots of the BMWP-Col with the others explanatory variables of the three models: BOD_5 (A); Log fecal coliforms (B); pH (C) and color (D).

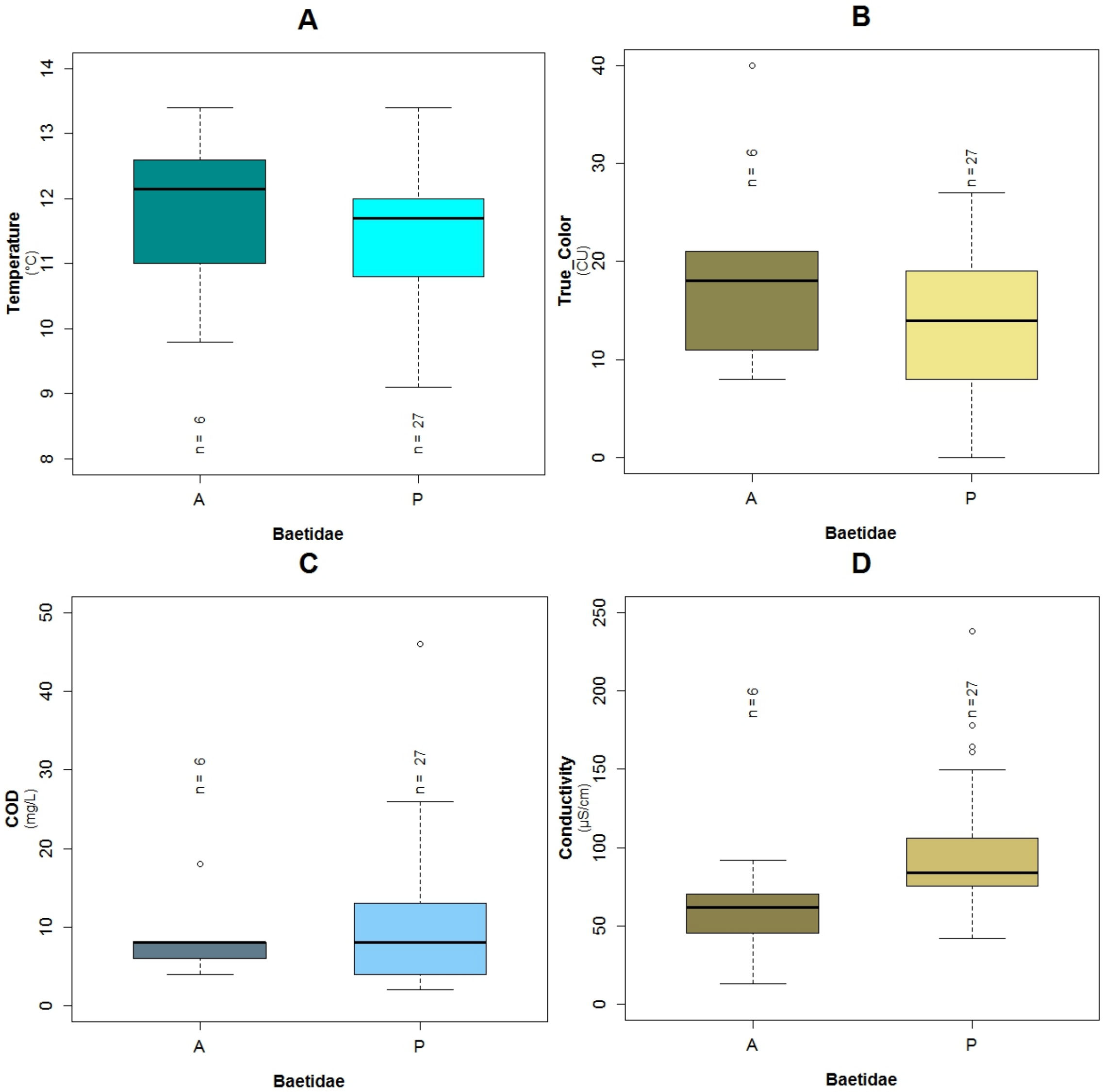


Fig. C4. Boxplots showing the presence (P) or absence (A) of *Baetidae* with its explanatory variables: (**A**) temperature; (**B**) color; (**C**) COD and (**D**) conductivity.

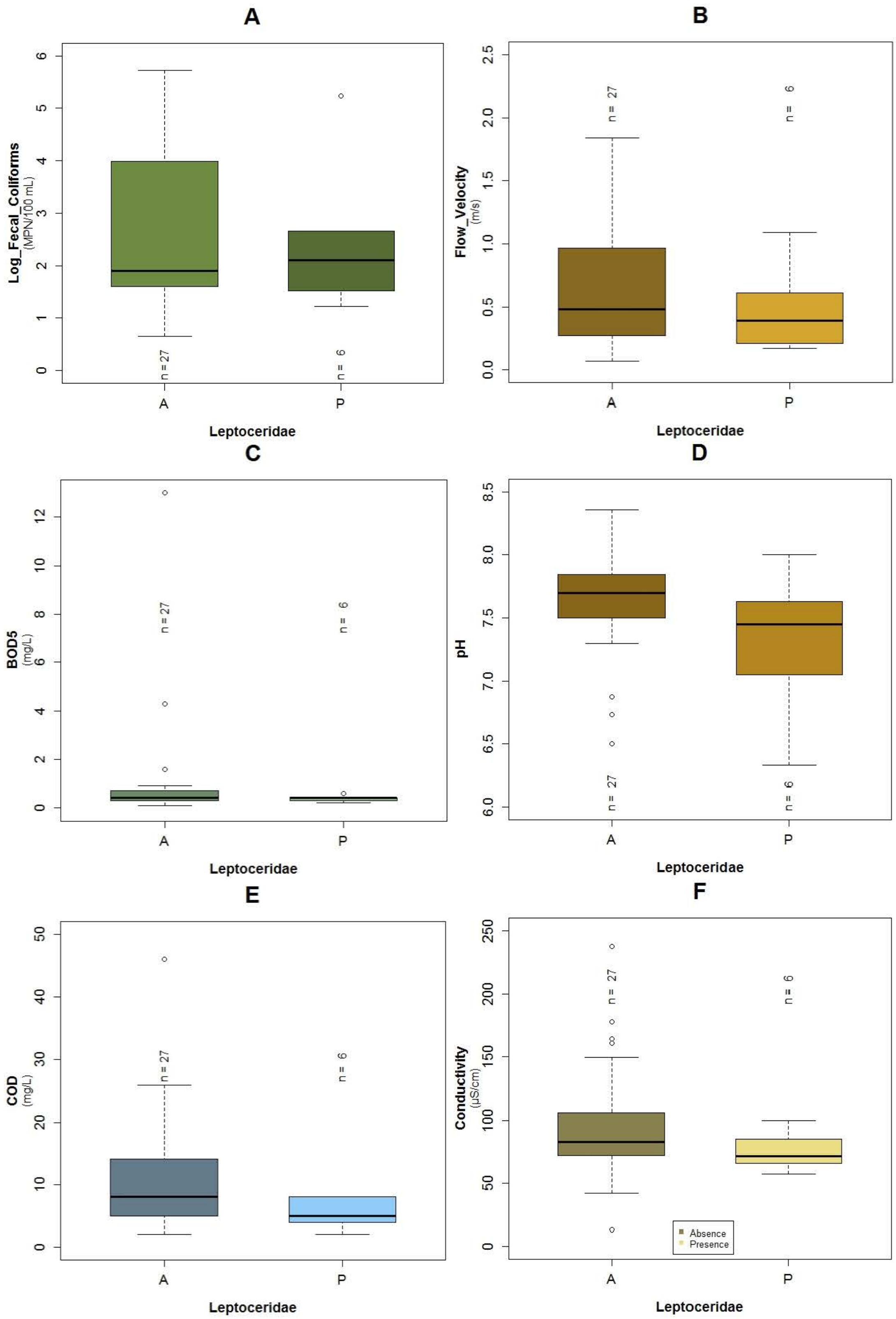


Fig. C5. Boxplots showing the presence (P) or absence (A) of *Leptoceridae* with its explanatory variables (**A**) fecal coliforms expressed in log scale; (**B**) flow velocity; (**C**) BOD_5; (**D**) pH; (**E**) COD and (**F**) conductivity.

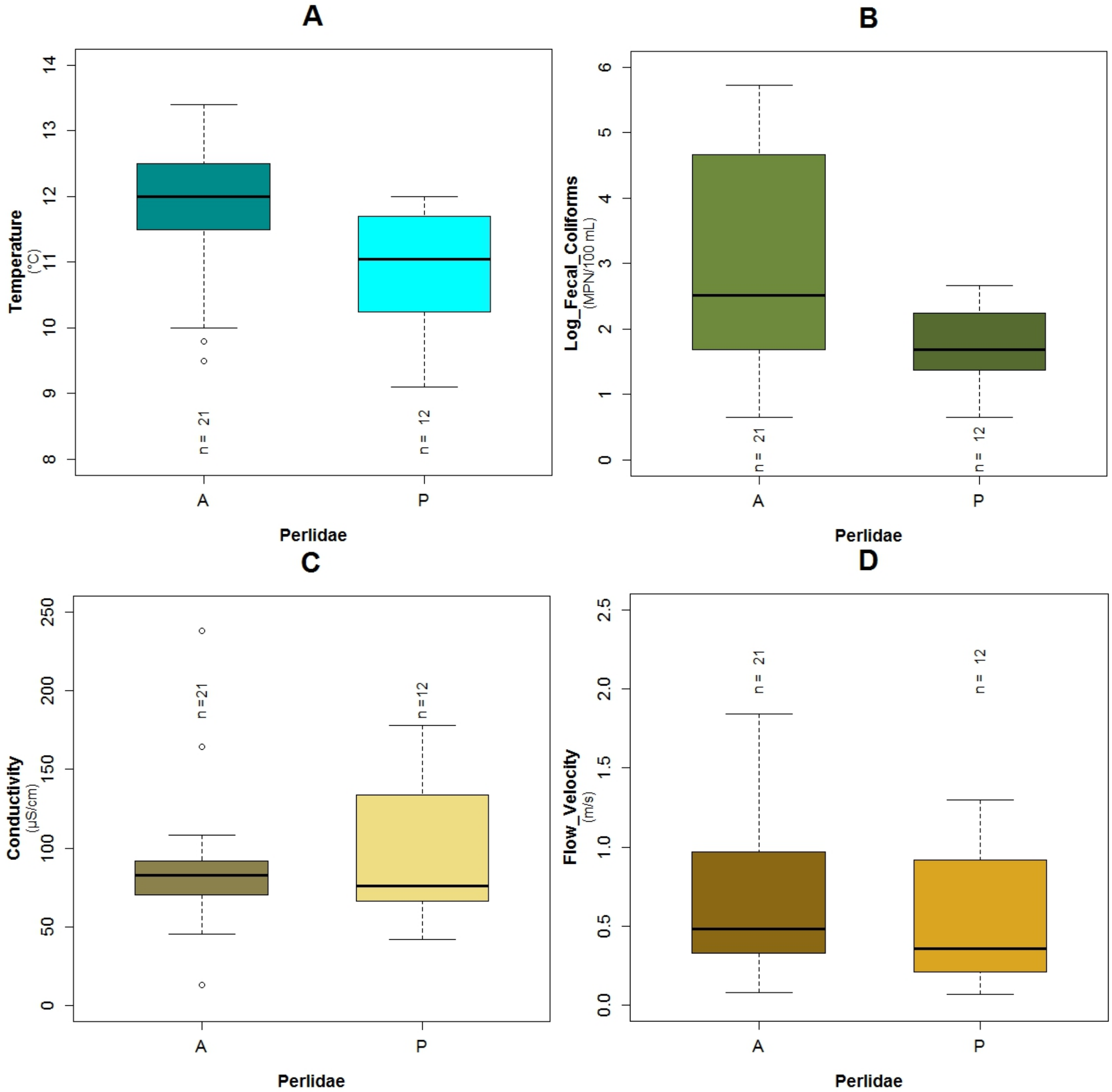


Fig. C6. Boxplots showing the presence or absence of *Perlidae* with its explanatory variables (**A**) temperature; (**B**) fecal coliforms expressed in log scale; (**C**) conductivity; (**D**) flow velocity.

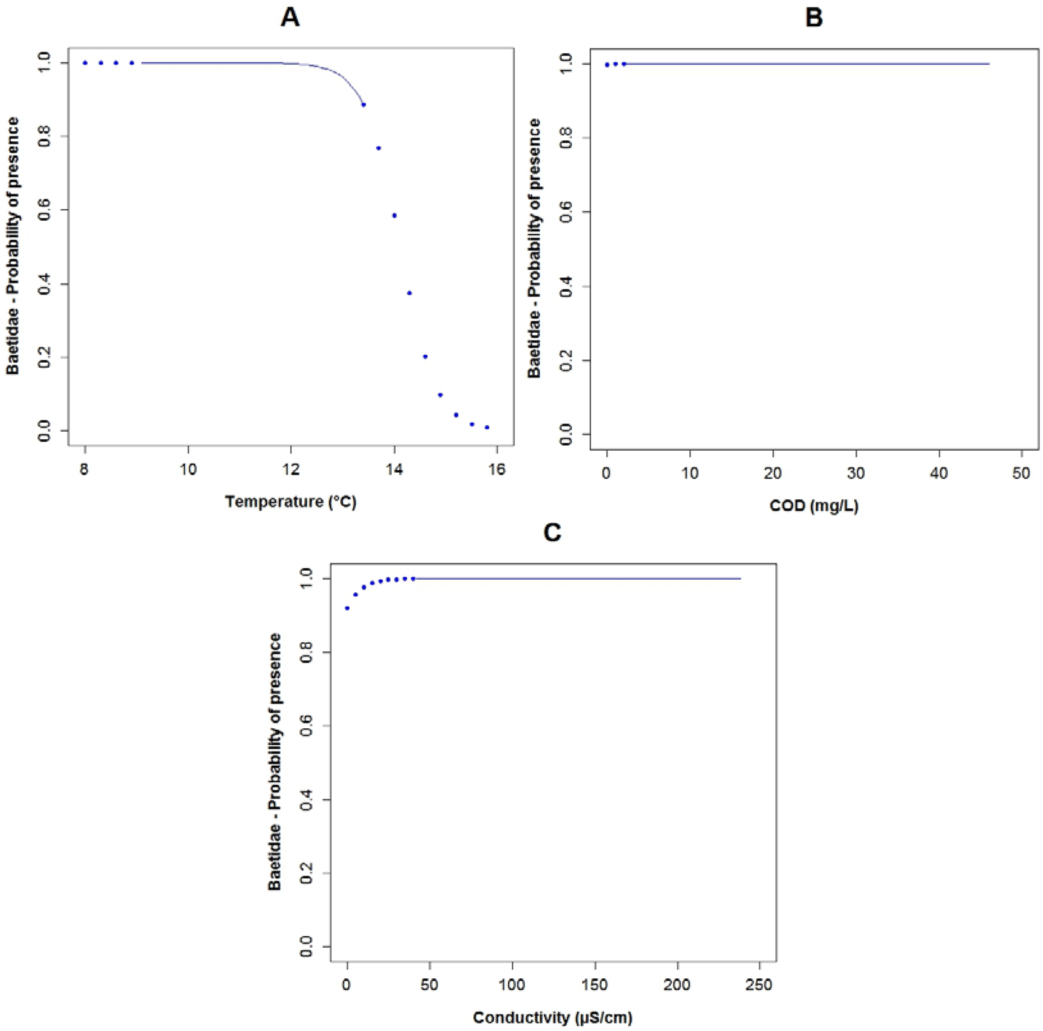


Fig. C7. The probability of *Baetidae* being present in relation to (**A**) temperature; (**B**) COD; (**C**) conductivity. (The blue points in the curves indicate extrapolation outside the observed physicochemical variables range).

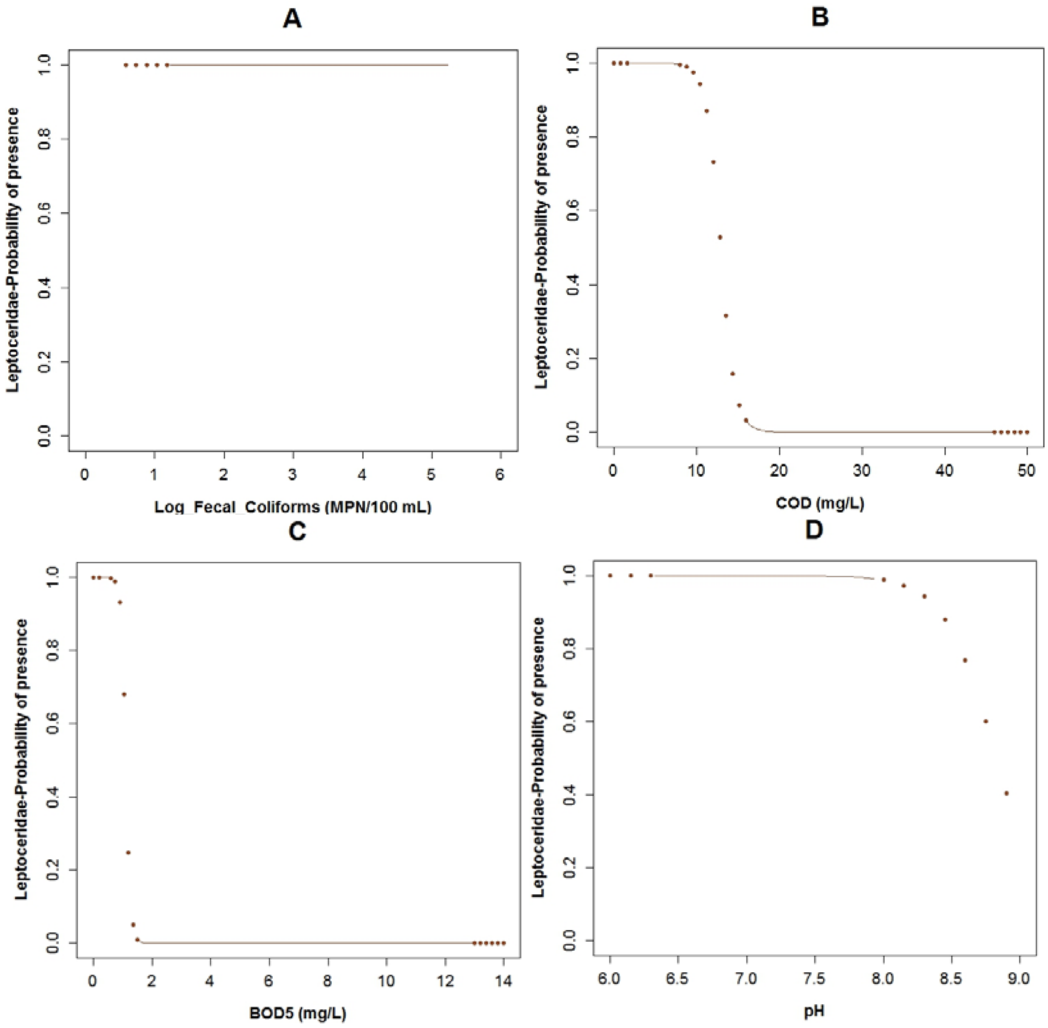


Fig. C8. The probability of *Leptoceridae* being present in relation to (**A**) log fecal coliforms; (**B**) COD; (**C**) BOD_5; (**D**) pH. (The brown points in the curves indicate extrapolation outside the observed physicochemical variables) range.

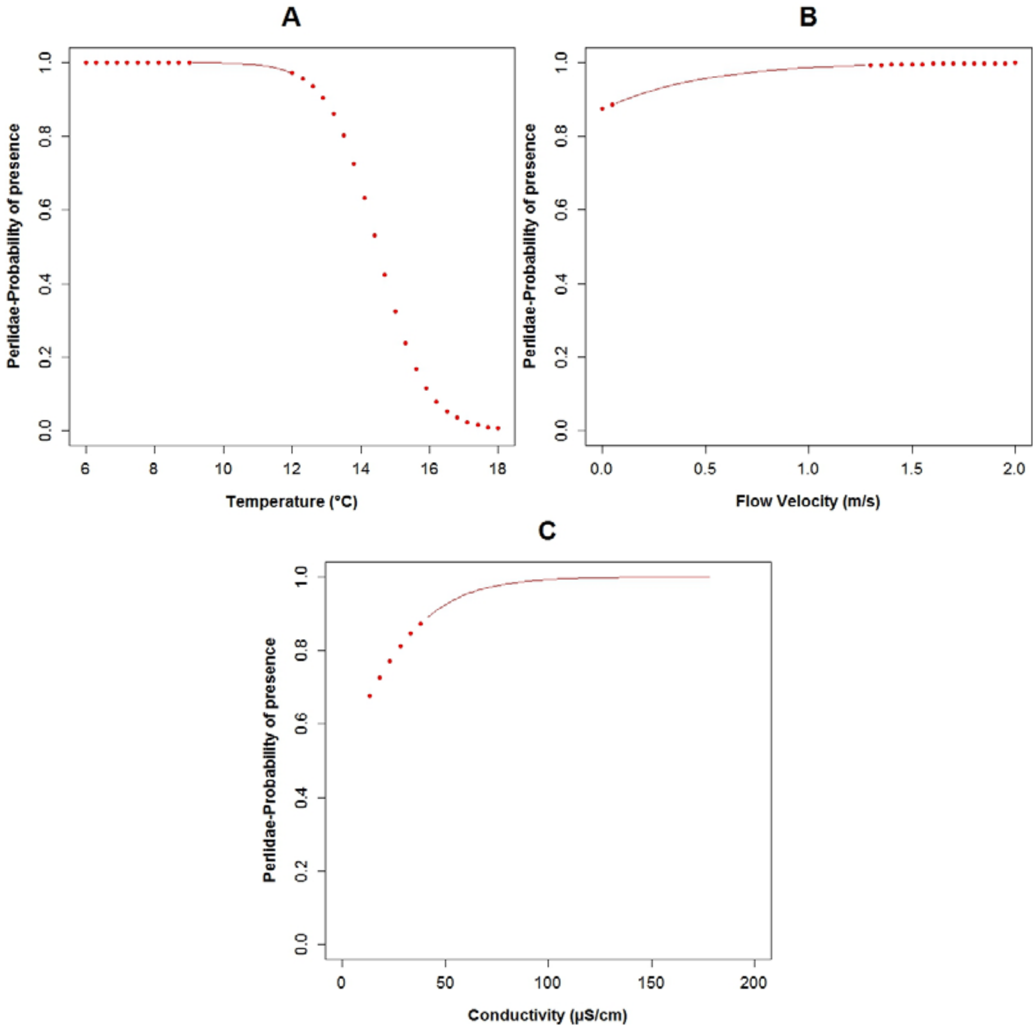


Fig. C9. The probability of *Perlidae* being present in relation to: (**A**) temperature; (**B**) flow velocity; (**C**) conductivity. (The red points in the curves indicate extrapolation outside the observed physicochemical variables range).

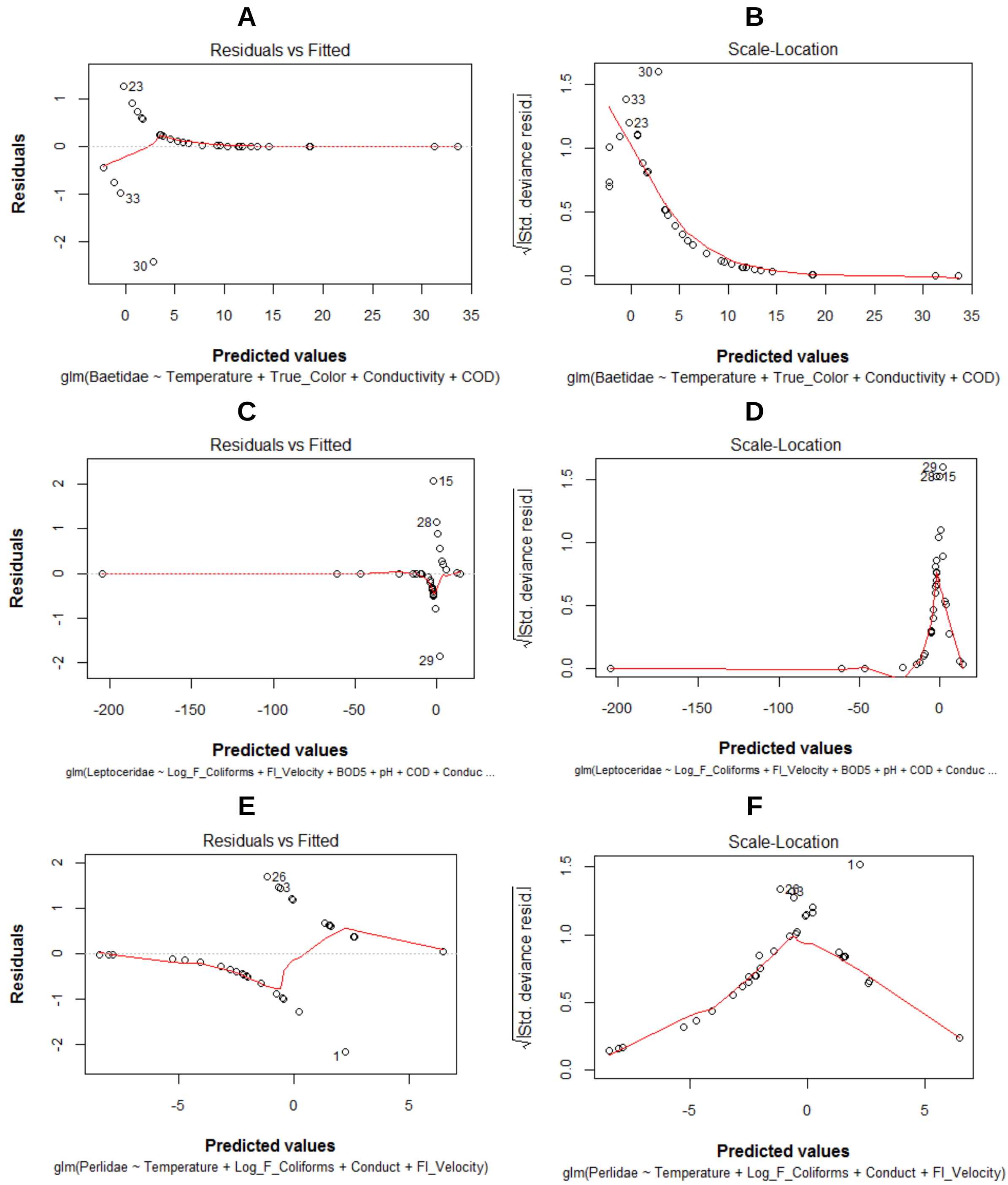


Fig. C10. Plots of Residual vs. Fitted (**A**) *Baetidae*; (**C**) *Leptoceridae*; (**E**) *Perlidae* and plots of Scale-Location of (**B**) *Baetidae*; (**D**) *Leptocerida* and (**F**) *Perlidae*.

Table C1. Spearman correlation *p*-values of the explanatory variable used to construct the generalized linear models (GLMs).

Explanatory Variable	BOD_5	COD	Conductivity	Flow Velocity	Log Fecal Coliforms	pH	True Color (Color)	Water Temperature
BOD_5								
COD	<0.01							
Conductivity	0.73	0.96						
Flow velocity	0.53	0.15	0.10					
Log fecal coliforms	<0.01	<0.01	0.04	0.30				
pH	0.23	0.40	0.38	0.50	0.02			
True color (color)	<0.01	<0.01	0.08	0.10	0.06	0.66		
Water temperature	<0.01	0.27	0.09	0.88	0.14	1.00	0.91	

Table C2. Regression parameters: Standard Error and *z* value of the models for predicting the presence of *Baetidae*, *Leptoceridae* and *Perlidae*.

Explanatory Variable	Baetidae		Leptoceridae		Perlidae	
	Std. Error	*z* Value	Std. Error	*z* Value	Std. Error	*z* Value
	16.9	1.6	22.0	2.3	6.5	2.2
BOD_5			6.6	−1.9		
COD	0.5	1.6	0.6	−2.0		
Conductivity	0.1	2.1	0.0	−1.9	0.0	2.1
Flow velocity			5.8	−2.1	1.7	1.4
Log fecal coliforms			3.1	2.1	0.8	−2.3
pH			2.4	−2.2		
Temperature	1.6	−1.8			0.6	−2.2
True color (color)	0.3	−1.7				

Table C3. Process of model development for *Baetidae*: *p*-values, AIC and pseudo R^2.

Model	(Intercept)	p-Value																AIC	R^2 (%)
		Flow Velocity	BOD_5	Turbidity	Mean Depth	Temperature	pH	DO	Color	Conductivity	COD	Nitrate. Nitrite	Ammonium Nitrogen	Organic Nitrogen	Total Solids	Log Fecal Coliforms	Phosphates		
m1301	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	40	100
m1302	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	38	100
m1303	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	-	36	100
m1304	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	-	34	100
m1305	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	-	32	100
m1306	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	-	30	100
m1307	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	-	28	100
m1308	1.00	1.00	1.00	1.00	-	1.00	1.00	1.00	1.00	1.00	1.00	-	-	1.00	1.00	1.00	-	26	100
m1309	1.00	-	1.00	1.00	-	1.00	1.00	1.00	1.00	1.00	1.00	-	-	1.00	1.00	1.00	-	24	100
m1310	1.00	-	1.00	1.00	-	1.00	-	1.00	1.00	1.00	1.00	-	-	1.00	1.00	1.00	-	22	100
m1311	1.00	-	1.00	1.00	-	1.00	-	1.00	1.00	1.00	1.00	-	-	-	1.00	1.00	-	20	100
m1312	0.97	-	0.98	0.97	-	0.97	-	0.97	0.97	0.97	0.97	-	-	-	0.97	-	-	18	100
m1313	0.65	-	-	0.74	-	0.65	-	0.67	0.70	0.97	0.70	-	-	-	0.70	-	-	24	75
m1314	0.31	-	-	0.41	-	0.34	-	0.31	0.38	-	0.42	-	-	-	0.33	-	-	22	75
m1315	0.08	-	-	0.18	-	0.10	-	0.08	0.12	-	-	-	-	-	0.34	-	-	26	56
m1316	0.04	-	-	0.09	-	0.08	-	0.06	0.05	-	-	-	-	-	-	-	-	25	52
m1317	0.04	-	-	-	-	0.11	-	0.06	0.06	-	-	-	-	-	-	-	-	31	26
m1318	0.04	-	-	-	-	-	-	0.16	0.11	-	-	-	-	-	-	-	-	33	15
m1319	0.09	-	-	-	-	-	-	0.20	0.00	-	-	-	-	-	-	-	-	34	6
m1320	0.08	-	-	0.26	-	0.08	-	0.14	0.25	0.27	-	-	-	-	-	-	-	25	58
m1321	0.05	-	-	0.08	-	0.07	-	0.10	0.07	-	0.30	-	-	-	-	-	-	25	57
m1322	0.07	-	0.85	0.12	-	0.14	-	0.07	0.07	-	-	-	-	-	-	-	-	27	53
m1323	0.21	-	-	0.20	-	0.06	0.39	0.13	0.08	-	-	-	-	-	-	-	-	26	55
m1324	0.98	-	-	0.98	-	0.98	-	0.98	0.98	-	-	-	-	-	-	0.98	-	12	100
m1325	0.99	-	-	0.11	-	0.08	-	0.08	0.06	-	-	-	-	-	-	-	1.00	27	54
m1326	0.05	-	-	0.10	-	0.09	-	0.06	0.05	-	-	1.0	-	-	-	-	-	27	52
m1327	0.05	-	-	0.10	-	0.09	-	0.06	0.05	-	-	-	1.00	-	-	-	-	27	53
m1328	0.05	-	-	0.10	-	0.09	-	0.06	0.06	-	-	0.83	-	-	-	-	-	27	53
m1329	0.09	-	-	0.30	-	0.11	-	0.11	0.26	-	-	-	-	0.33	-	-	-	25	58
m1330	0.08	-	-	0.18	-	0.10	-	0.08	0.12	-	-	-	-	-	0.34	-	-	26	56
m1331	1.00	-	-	1.00	-	1.00	-	1.00	1.00	-	-	-	-	-	-	-	-	12	100
m1332	0.99	-	-	0.11	-	0.08	-	0.08	0.06	-	-	-	-	-	-	-	-	27	54
m1333	0.05	-	-	0.07	0.23	0.07	-	0.08	0.07	-	-	-	-	-	-	-	-	25	59
m1334	0.15	0.12	-	0.13	-	0.17	-	0.15	0.07	-	-	-	-	-	-	-	-	21	72
m1335	0.12	0.29	-	0.21	-	0.00	-	0.18	0.07	-	-	-	-	-	-	-	-	30	38
m1336	0.07	-	-	0.15	-	0.00	-	0.17	0.07	-	-	-	-	-	-	-	-	29	33
m1337	1.00	1.00	-	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	-	-	1.00	1.00	-	-	24	100
m1338	1.00	1.00	-	1.00	-	1.00	1.00	1.00	1.00	1.00	1.00	-	-	1.00	1.00	-	-	22	100
m1339	1.00	1.00	-	1.00	-	1.00	-	1.00	1.00	1.00	1.00	-	-	1.00	1.00	-	-	20	100
m1340	0.99	-	-	0.99	-	0.99	-	0.99	0.99	0.99	0.99	-	-	0.99	1.00	-	-	18	100
m1341	0.99	-	-	0.99	-	0.99	-	0.99	0.99	0.99	0.99	-	-	0.99	-	-	-	16	100
m1342	0.10	-	-	0.35	-	0.12	-	0.40	0.17	0.17	0.31	-	-	-	-	-	-	24	67
m1343	0.14	-	-	0.36	-	0.12	-	-	0.14	0.09	0.20	-	-	-	-	-	-	23	65
m1344	0.11	-	-	-	-	0.07	-	-	0.10	0.04	0.10	-	-	-	-	-	-	22	60

Table C4. Process of model development for *Leptoceridae*: *p*-values, AIC and pseudo R^2.

Model	(Intercept)	*p*-Value																AIC	R^2 (%)
		Flow Velocity	BOD_5	Turbidity	Mean Depth	Temperature	pH	DO	Color	Conductivity	COD	Nitrate. Nitrite	Ammonium Nitrogen	Organic Nitrogen	Total Solids	Log Fecal Coliforms	Phosphates		
m15	0.69	-	-	-	0.05	-	-	-	-	0.80	-	0.11	-	0.50	-	0.11	-	45	25
m150	0.82	-	-	-	0.05	-	-	-	-	-	-	0.11	-	-	-	0.11	-	37	24
m152	0.30	-	-	-	0.07	-	-	-	-	-	-	0.26	-	-	-	-	-	38	16
m153	0.86	-	-	-	0.13	-	-	-	-	-	-	-	-	-	-	0.65	-	41	10
m154	0.86	-	-	-	0.14	-	-	-	-	-	-	-	-	-	-	-	-	39	9
m155	0.51	0.12	-	-	0.15	-	-	-	-	-	-	0.07	-	-	-	0.07	-	36	32
m156	0.88	0.04	-	-	-	-	-	-	-	-	-	0.11	-	-	-	0.08	-	38	23
m157	0.63	0.17	-	-	-	-	-	-	-	-	-	-	-	-	-	0.66	-	42	6
m158	0.52	0.04	0.10	-	-	-	-	-	-	-	-	0.12	-	-	-	0.03	-	34	38
m159	0.89	0.12	0.07	-	-	-	-	-	-	-	-	-	-	-	-	0.04	-	37	26
m1590	1.00	0.24	0.10	-	0.69	-	-	-	-	-	-	-	-	-	-	0.05	-	39	26
m1591	0.93	0.10	0.06	0.19	-	-	-	-	-	-	-	-	-	-	-	0.02	-	36	32
m1592	1.00	0.07	0.22	0.19	-	-	-	-	-	-	-	-	-	-	-	0.03	-	37	36
m1593	0.63	0.08	0.05	0.26	-	-	-	-	-	0.55	-	-	-	-	-	0.02	-	38	33
m1594	0.85	0.10	0.06	0.20	-	-	-	0.87	-	-	-	-	-	-	-	0.03	-	38	32
m1595	0.95	0.11	0.07	0.20	-	0.97	-	-	-	-	-	-	-	-	-	0.02	-	38	32
m1596	0.71	0.10	0.05	0.25	-	-	-	-	-	-	-	-	-	-	0.63	0.02	-	38	32
m1597	0.17	0.08	0.06	0.48	-	-	0.17	-	-	-	-	-	-	-	-	0.02	-	36	38
m1598	0.08	0.07	0.06	-	-	-	0.07	-	-	-	-	-	-	-	-	0.03	-	35	36
m1599	0.04	0.05	0.15	-	-	-	0.04	-	-	-	0.11	-	-	-	-	0.02	-	32	48
m1580	0.05	0.12	-	-	-	-	0.06	-	-	-	0.03	-	-	-	-	0.04	-	35	36
m1581	0.13	-	-	-	-	-	0.12	-	-	-	0.03	-	-	-	-	0.05	-	36	28
m1582	0.75	-	-	-	-	-	-	-	-	-	0.04	-	-	-	-	0.09	-	37	20
m1583	0.04	0.09	0.16	-	0.60	-	0.04	-	-	-	0.10	-	-	-	-	0.02	-	34	49
m1584	0.02	0.03	0.06	-	-	-	0.03	-	-	0.06	0.05	-	-	-	-	0.04	-	27	68
m1585	0.07	0.04	0.05	0.57	-	-	0.09	-	-	0.07	0.11	-	-	-	-	0.04	-	28	69
m1586	0.04	0.03	0.07	-	-	-	0.05	-	-	0.09	0.09	0.90	-	-	-	0.05	-	29	68
m1575	0.02	0.03	0.07	-	0.43	-	0.03	-	-	0.08	0.09	-	-	-	-	0.05	-	28	70
m1587	0.04	0.12	0.13	-	-	0.63	0.05	-	-	0.15	0.08	-	-	-	-	0.10	-	28	68
m1588	0.02	0.03	0.05	-	-	-	0.12	0.61	-	0.06	0.10	-	-	-	-	0.03	-	28	68
m1589	0.97	0.05	0.20	-	-	-	0.04	-	-	0.07	0.07	-	-	-	-	0.05	-	28	69
m1570	0.06	0.06	0.08	-	-	-	0.07	-	-	0.09	0.12	-	-	-	0.36	0.07	-	27	71
m1571	0.03	0.04	0.08	-	-	-	0.04	-	0.45	0.07	0.09	-	-	-	-	0.04	-	29	69
m1572	0.03	0.04	0.16	-	-	-	0.04	-	-	0.06	0.06	-	1.00	-	-	0.04	-	28	68
m1573	0.04	0.04	0.09	-	-	-	0.05	-	-	0.10	0.09	0.75	-	-	-	0.05	-	28	68
m1574	0.97	0.05	0.20	-	-	-	0.04	-	-	0.07	0.07	-	-	-	-	0.05	1.00	28	69

Table C5. Process of model development for *Perlidae*: *p*-values, AIC and pseudo R^2.

Model	(Intercept)	p-Value																AIC	R^2 (%)
		Flow Velocity	BOD_5	Turbidity	Mean Depth	Temperature	pH	DO	Color	Conductivity	COD	Nitrate. Nitrite	Ammonium Nitrogen	Organic Nitrogen	Total Solids	Log Fecal Coliforms	Phosphates		
m10	0.90	-	-	0.31	-	0.06	0.38	0.38	0.24	0.40	0.51	0.10	0.98	0.29	-	0.45	-	39	66
m11	0.90	-	-	0.31	-	0.06	0.38	0.38	0.24	0.40	0.51	0.10	-	0.29	-	0.45	-	37	66
m12	0.84	-	-	0.21	-	0.04	0.49	0.53	0.24	0.44	-	0.09	-	0.17	-	0.39	-	35	65
m13	0.79	-	-	0.22	-	0.05	0.63	-	0.21	0.42	-	0.10	-	0.17	-	0.53	-	34	64
m14	0.07	-	0.51	0.20	-	0.03	-	-	0.32	0.49	-	0.18	-	0.20	-	0.31	-	34	64
m15	0.02	-	-	0.88	-	0.02	-	-	0.42	-	-	0.52	-	0.49	-	-	-	43	29
m16	0.05	0.75	0.57	-	0.88	0.07	-	-	0.68	-	-	-	-	0.46	-	-	-	44	31
m160	0.05	0.78	0.54	-	-	0.07	-	-	0.70	-	-	-	-	0.45	-	-	-	42	31
m161	0.05	-	0.53	-	-	0.07	-	-	0.75	-	-	-	-	0.43	-	-	-	40	31
m162	0.05	-	0.43	-	-	0.07	-	-	-	-	-	-	-	0.29	-	-	-	38	30
m163	0.02	-	-	-	-	0.02	-	-	-	-	-	-	-	0.23	-	-	-	39	24
m164	0.52	-	-	-	-	-	-	-	-	-	-	-	-	0.33	-	-	-	46	4
m165	0.05	-	-	-	-	0.07	-	0.52	-	-	-	-	-	-	-	0.08	-	40	26
m166	0.02	-	-	0.52	-	0.02	-	-	-	-	-	-	-	0.23	-	0.00	-	41	25
m167	0.03	-	-	-	-	0.05	-	-	-	-	-	-	-	0.27	-	0.11	-	37	32
m168	0.03	-	-	-	-	0.04	-	0.51	-	-	-	-	-	0.26	-	0.10	-	39	33
m169	0.16	0.83	0.06	-	-	-	-	-	-	-	-	-	-	-	-	0.00	-	41	19
m170	0.14	-	0.06	-	-	-	-	-	-	-	-	-	-	-	-	0.00	-	39	19
m171	0.08	-	0.20	-	-	0.14	-	-	-	-	-	-	-	-	-	0.00	-	39	25
m172	0.10	-	0.21	-	-	0.00	-	-	-	-	-	-	-	-	-	0.32	-	40	22
m173	0.07	-	0.48	-	-	0.15	-	-	-	-	-	-	-	-	-	0.34	-	40	27
m174	0.04	-	0.72	-	-	0.07	-	-	-	-	-	-	-	0.28	-	0.32	-	39	33
m175	0.03	-	-	-	-	0.02	-	-	-	-	-	-	-	-	-	-	-	41	15
m176	0.99	-	-	0.10	-	-	-	-	-	-	0.17	-	-	-	-	0.09	1.00	40	30
m177	0.13	-	-	0.18	-	-	-	-	-	-	0.14	-	-	-	0.56	0.09	-	41	29
m178	0.42	-	-	0.17	-	-	-	-	-	0.28	0.22	-	-	-	-	0.07	-	40	31
m179	0.79	-	-	0.13	-	-	0.90	-	-	0.00	0.09	-	-	-	-	0.09	-	41	28
m1601	0.04	-	-	0.11	-	-	-	-	-	0.00	0.08	-	-	-	-	0.09	-	39	28
m1602	0.03	-	-	0.12	-	0.08	-	-	-	0.00	0.07	-	-	-	-	0.16	-	37	37
m1603	0.03	-	-	0.38	-	0.04	-	-	-	0.00	0.04	-	-	-	-	-	-	39	29
m1604	0.03	-	-	-	-	0.05	-	-	-	0.00	0.06	-	-	-	-	-	-	38	27
m1605	0.04	-	-	-	-	0.08	-	-	-	0.00	0.15	-	-	-	-	0.28	-	38	31
m1606	0.03	-	-	-	-	0.06	-	0.24	-	0.00	0.10	-	-	-	-	0.23	-	39	34
m1607	0.03	-	-	-	-	0.04	-	0.32	-	0.00	0.04	-	-	-	-	-	-	38	30
m1608	0.20	-	-	-	-	0.05	0.98	-	-	0.00	0.06	-	-	-	-	-	-	40	27
m1609	0.03	-	-	-	-	0.04	-	-	-	0.00	0.08	0.52	-	-	-	-	-	39	29
m1610	0.99	-	-	-	-	0.06	-	-	-	0.00	0.12	-	-	-	-	-	1.00	39	29
m1611	0.04	-	-	-	-	0.04	-	-	-	0.22	0.06	-	-	-	-	-	-	38	31
m1612	0.03	0.54	-	-	-	0.04	-	-	-	-	0.05	-	-	-	-	-	-	39	28
m1613	0.03	-	-	-	0.89	0.06	-	-	-	-	0.06	-	-	-	-	-	-	39	27
m1614	0.04	-	-	-	-	0.04	-	-	-	-	0.09	-	-	-	0.60	-	-	39	28
m1615	0.03	-	-	-	-	0.04	-	-	0.54	-	0.28	-	-	-	-	-	-	39	28
m1615'	0.06	-	0.51	-	-	0.12	-	-	-	-	0.19	-	-	-	-	-	-	39	29
m1616	0.04	0.23	0.96	0.22	-	0.03	-	0.22	-	0.19	0.25	0.88	-	-	-	0.12	-	40	54
m1617	0.03	0.21	-	0.19	-	0.03	-	0.19	-	0.18	0.21	0.89	-	-	-	0.09	-	38	54
m1618	0.03	0.14	-	0.18	-	0.03	-	0.18	-	0.17	0.19	-	-	-	-	0.07	-	36	53
m1619	0.04	0.25	-	0.66	-	0.03	-	0.57	-	0.04	-	-	-	-	-	0.02	-	38	45
m1620	0.04	0.16	-	-	-	0.02	-	0.59	-	0.04	-	-	-	-	-	0.02	-	36	44
m1621	0.03	0.17	-	-	-	0.02	-	-	-	0.04	-	-	-	-	-	0.02	-	35	44
m1622	0.04	-	-	-	-	0.04	-	-	-	0.06	-	-	-	-	-	0.04	-	35	38
m1623	0.04	-	-	-	-	0.08	-	-	-	-	-	-	-	-	-	0.09	-	38	25
m1624	0.14	-	0.06	-	-	-	-	-	-	-	-	-	-	-	-	-	-	39	19
m1625	0.12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.04	-	40	17

Appendix D – Supporting information for chapter 5

Table D1. Models settings used in Weka before and after optimization analysis, in which was included Cost Matrix Weights (CMW).

Model No.	FCR [a]	Dataset macroinvertebrates	Model Settings					
			J4.8	PCF [b]	CMW [c]			
					TP [d]	FN [e]	FP [f]	TN [g]
* 1[h]ap[i]1[j]	Recreational	Presence/absence	3, 5 and 10 fcv [k]	0.25				
* 1ap2	Recreational	Presence/absence	3, 5 and 10 fcv	0.10				
1ap3	Recreational	Presence/absence	3, 5 and 10 fcv	0.25	0	1	2	0
1ap4	Recreational	Presence/absence	3, 5 and 10 fcv	0.25	0	1	3	0
1ap5	Recreational	Presence/absence	3, 5 and 10 fcv	0.25	0	1	5	0
1ap6	Recreational	Presence/absence	3, 5 and 10 fcv	0.25	0	1	6	0
1ap7	Recreational	Presence/absence	3, 5 and 10 fcv	0.25	0	1	7	0
1ap8	Recreational	Presence/absence	3, 5 and 10 fcv	0.10	0	1	10	0
* 1a1	Recreational	Abundance	3, 5, 10 fcv and 66%tr	0.25				
* 1a2	Recreational	Abundance	3, 5, 10 fcv and 66%tr	0.10				
1a3	Recreational	Abundance	3, 5 and 10 fcv	0.25	0	1	1	0
1a4	Recreational	Abundance	3, 5 and 10 fcv	0.25	0	1	2	0
1a5	Recreational	Abundance	3, 5 and 10 fcv	0.25	0	1	3	0
1a6	Recreational	Abundance	3, 5 and 10 fcv	0.25	0	1	4	0
1a7	Recreational	Abundance	3, 5 and 10 fcv	0.25	0	1	5	0
1a8	Recreational	Abundance	3, 5 and 10 fcv	0.25	0	1	7	0
1a9	Recreational	Abundance	3, 5 and 10 fcv	0.25	0	1	8	0
1a10	Recreational	Abundance	3, 5 and 10 fcv	0.1	0	1	8	0
1a11	Recreational	Abundance	3, 5 and 10 fcv	0.25	0	1	9	0
1a12	Recreational	Abundance	3, 5 and 10 fcv	0.1	0	1	9	0
1a13	Recreational	Abundance	3, 5 and 10 fcv	0.25	0	1	10	0
1a14	Recreational	Abundance	3, 5 and 10 fcv	0.1	0	1	10	0
1a15	Recreational	Abundance	3, 5 and 10 fcv	0.25	0	1	11	0
1a16	Recreational	Abundance	3, 5 and 10 fcv	0.25	0	1	12	0
1a17	Recreational	Abundance	3, 5 and 10 fcv	0.25	0	1	15	0
1a18	Recreational	Abundance	3, 5 and 10 fcv	0.25	0	1	18	0
* 2ap1	Agriculture l	Presence/absence	3, 5 and 10 fcv	0.25				
* 2ap2	Agriculture	Presence/absence	3, 5 and 10 fcv	0.10				
2ap3	Agriculture	Presence/absence	3, 5 and 10 fcv	0.25	0	1	2	0
2ap4	Agriculture	Presence/absence	3, 5 and 10 fcv	0.25	0	1	3	0
2ap5	Agriculture	Presence/absence	3, 5 and 10 fcv	0.25	0	1	5	0
* 2a1	Agriculture	Abundance	3, 5, 10 fcv and 66%tr	0.25				
* 2a2	Agriculture	Abundance	3, 5, 10 fcv and 66%tr	0.10				
2a3	Agriculture	Abundance	3, 5 and 10 fcv	0.25	0	1	1	0
2a4	Agriculture	Abundance	3, 5 and 10 fcv	0.25	0	1	2	0
2a5	Agriculture	Abundance	3, 5 and 10 fcv	0.25	0	1	3	0
2a6	Agriculture	Abundance	3, 5 and 10 fcv	0.25	0	1	4	0
2a7	Agriculture	Abundance	3, 5 and 10 fcv	0.25	0	1	5	0
2a8	Agriculture	Abundance	3, 5 and 10 fcv	0.25	0	1	8	0
2a9	Agriculture	Abundance	3, 5 and 10 fcv	0.25	0	1	10	0
2a10	Agriculture	Abundance	3, 5 and 10 fcv	0.25	0	1	12	0
2a11	Agriculture	Abundance	3, 5 and 10 fcv	0.25	0	1	15	0
2a12	Agriculture	Abundance	3, 5 and 10 fcv	0.25	0	1	17	0
2a13	Agriculture	Abundance	3, 5 and 10 fcv	0.25	0	1	18	0
2a14	Agriculture	Abundance	3, 5 and 10 fcv	0.25	0	1	20	0
2a15	Agriculture	Abundance	3, 5 and 10 fcv	0.25	0	1	21	0
2a16	Agriculture	Abundance	3, 5 and 10 fcv	0.25	0	1	22	0
2a17	Agriculture	Abundance	3, 5 and 10 fcv	0.25	0	1	25	0

* Model developed before optimization process, in which the CMW was not used. [a] FCR = Fecal coliform regulation. [b] PCF = Pruning confidence factor. [c] CMW = Cost Matrix Weights. [d] TP = True positives. [e] FN = False negative. [f] FP = False positive. [g] TN = True negative. [h] The number of FCR. [i] The kind of database: ap = absence/presence, a = abundance. [j] The number of model with different values of PCF and CMW. [k] fcv = folds cross validation. [l] Models obtained from agriculture could be applied to check raw water regulations.

Table D2. Predictable results of the models before and after the optimization process: Correctly classified instances (CCI), Kappa statistics, and overall confusion entropy of a confusion matrix (CEN).

Model No.	FCR [a]	Model Outcomes: CCI [b] (%)	Kappa Statistics	Number of Leaves	CEN [c]
		CCI ± sd	Kappa ± sd		CEN ± sd
* 1[e]ap[f]1[g]	Recreational [d]	40.4 ± 3.5	-0.2 ± 0.1	6	1.0 ± 0.0
* 1ap2	Recreational	48.5 ± 3.0	-0.1 ± 0.1	2	1.0 ± 0.0
1apf3[g]	Recreational	42.4 ± 6.1	-0.1 ± 0.1	3	1.0 ± 0.0
1ap4	Recreational	41.4 ± 7.6	-0.1 ± 0.2	4	0.9 ± 0.1
1ap5	Recreational	43.4 ± 1.7	0.0 ± 0.0	4	0.8 ± 0.1
1ap6	Recreational	42.4 ± 3.0	0.0 ± 0.0	4	0.8 ± 0.1
1ap7	Recreational	42.4 ± 3.0	0.0 ± 0.1	1	0.8 ± 0.1
1ap8	Recreational	42.4 ± 0.0	0.0 ± 0.0	1	0.6 ± 0.1
* 1a1	Recreational	70.5 ± 1.5	0.4 ± 0.0	5	0.8 ± 0.0
* 1a2	Recreational	70.5 ± 1.5	0.4 ± 0.0	4	0.8 ± 0.0
1a3	Recreational	69.7 ± 0.0	0.4 ± 0.0	5	0.8 ± 0.0
1a4	Recreational	72.7 ± 6.1	0.4 ± 0.1	4	0.8 ± 0.1
1a5	Recreational	77.8 ± 4.6	0.6 ± 0.1	3	0.6 ± 0.1
1a6	Recreational	76.8 ± 9.2	0.5 ± 0.2	3	0.6 ± 0.1
1a7	Recreational	77.7 ± 12.2	0.6 ± 0.2	3	0.6 ± 0.2
1a8	Recreational	78.8 ± 8.0	0.6 ± 0.2	3	0.6 ± 0.1
1a9	Recreational	79.8 ± 1.7	0.6 ± 0.0	3	0.6 ± 0.1
1a10	Recreational	79.8 ± 1.7	0.6 ± 0.0	3	0.6 ± 0.1
1a11	Recreational	70.7 ± 7.6	0.4 ± 0.1	3	0.7 ± 0.1
1a12	Recreational	70.7 ± 7.6	0.4 ± 0.1	3	0.7 ± 0.1
1a13	Recreational	66.7 ± 10.9	0.4 ± 0.2	3	0.7 ± 0.1
1a14	Recreational	63.6 ± 12.1	0.3 ± 0.2	3	0.7 ± 0.0
1a15	Recreational	62.6 ± 10.6	0.3 ± 0.2	3	0.7 ± 0.0
1a16	Recreational	58.6 ± 6.3	0.2 ± 0.1	3	0.7 ± 0.0
1a17	Recreational	47.5 ± 8.8	0.0 ± 0.0	3	0.6 ± 0.1
1a18	Recreational	42.4 ± 0.0	0.0 ± 0.0	1	0.5 ± 0.0
* 2ap1	Agriculture	66.7 ± 0.0	0.2 ± 0.0	4	0.9 ± 0.0
* 2ap2	Agriculture	69.7 ± 3.0	0.2 ± 0.0	3	0.8 ± 0.0
2ap3	Agriculture	75.8 ± 5.2	0.5 ± 0.1	4	0.7 ± 0.1
2ap4	Agriculture	73.7 ± 6.3	0.4 ± 0.1	3	0.7 ± 0.1
2ap5	Agriculture	71.7 ± 9.3	0.4 ± 0.2	3	0.7 ± 0.1
* 2a1	Agriculture	86.4 ± 8.0	0.7 ± 0.2	3	0.5 ± 0.2
* 2a2	Agriculture	77.3 ± 16.9	0.4 ± 0.4	3	0.7 ± 0.3
2a3	Agriculture	84.9 ± 9.1	0.6 ± 0.2	3	0.6 ± 0.2
2a4	Agriculture	89.9 ± 4.6	0.7 ± 0.1	3	0.5 ± 0.1
2a5	Agriculture	86.9 ± 7.6	0.7 ± 0.2	3	0.5 ± 0.2
2a6	Agriculture	86.9 ± 7.6	0.7 ± 0.2	3	0.5 ± 0.2
2a7	Agriculture	85.9 ± 6.3	0.7 ± 0.2	3	0.6 ± 0.1
2a8	Agriculture	80.8 ± 9.8	0.6 ± 0.2	3	0.6 ± 0.1
2a9	Agriculture	80.8 ± 9.7	0.6 ± 0.2	3	0.6 ± 0.1
2a10	Agriculture	80.8 ± 15.0	0.6 ± 0.3	3	0.6 ± 0.2
2a11	Agriculture	72.7 ± 6.1	0.4 ± 0.1	3	0.7 ± 0.1
2a12	Agriculture	60.6 ± 9.1	0.3 ± 0.1	3	0.7 ± 0.1
2a13	Agriculture	63.6 ± 10.5	0.3 ± 0.1	3	0.7 ± 0.0
2a14	Agriculture	57.6 ± 5.3	0.2 ± 0.1	2	0.7 ± 0.0
2a15	Agriculture	58.6 ± 7.0	0.3 ± 0.1	2	0.7 ± 0.0
2a16	Agriculture	47.5 ± 19.7	0.2 ± 0.2	2	0.6 ± 0.1
2a17	Agriculture	40.4 ± 13.6	0.1 ± 0.1	2	0.6 ± 0.1

* Model developed before optimization process, in which the CMW was not used. Mean and standard deviations of CCI, Kappa statistics and CEN were derived from threefold cross validation. [a] FCR = Fecal coliform regulation. [b] CCI = Correctly classified instances. [c] CEN = Overall confusion entropy of a confusion matrix. [d] The short name of FCR. [e] 1 is used for recreational coliforms regulation, while 2 for agriculture coliforms regulation. [f] The kind of database: ap = absence/presence, a = abundance. [g] The number of model with different value of PCF.

Table D3. Representation of the decision tree models (DTMs) in relation to Water Use Standard fulfillment. (a) Recreational regulation: primary contact (b) Agriculture regulation: agriculture and livestock irrigation.

<table>
<tr><td colspan="2">(a) Recreational Fecal Coliform Regulation</td></tr>
<tr><td>Model: 1a4</td><td>Models: 1a5, 1a6, 1a7</td></tr>
<tr><td>Baetidae <= 3: A
Baetidae > 3
| Perlidae = 0: B
| Perlidae > 0
| | Chironomidae <= 3: B
| | Chironomidae > 3: A</td><td>Baetidae <= 3
| Scirtidae = 1: A
| Scirtidae > 1: B
Baetidae > 3: B</td></tr>
<tr><td>Model: 1a8</td><td>Models: 1a9, 1a10, 1a11, 1a12</td></tr>
<tr><td>Baetidae <= 3
| Elminthidae <= 2: A
| Elminthidae > 2: B
Baetidae > 3: B</td><td>Baetidae <= 3
| Scirtidae <= 4: A
| Scirtidae > 4: B
Baetidae > 3: B</td></tr>
<tr><td colspan="2">(b) Agriculture fecal coliform regulation</td></tr>
<tr><td>Model: 2ap3</td><td>Models: 2ap4, 2ap5</td></tr>
<tr><td>Perlidae = presence: A
Perlidae = absence
| Baetidae = presence
| | Leptophlebiidae = presence: A
| | Leptophlebiidae = absence: B
| Baetidae = absence: A</td><td>Perlidae = presence: A
Perlidae = absence
| Baetidae = presence: B
| Baetidae = absence: A</td></tr>
<tr><td>Models: 2a1, 2a2, 2a3, 2a4, 2a5, 2a6</td><td>Models: 2a7, 2a8, 2a9, 2a10, 2a11</td></tr>
<tr><td>Baetidae <= 4: A
Baetidae > 4
| Perlidae = 0: B
| Perlidae > 0: A</td><td>Perlidae = 0
| Baetidae <= 4: A
| Baetidae > 4: B
Perlidae > 0: A</td></tr>
</table>

A: fulfillment; B: non-fulfillment

Table D4. Verification of the fulfilment of the recreational with primary contact Ecuadorian water use regulations associated with fecal coliforms according the decision tree models (DTMs)

Sampled Points	Land Use	Recreational fecal coliform regulation (threshold = 2.00E+02 MPN/100 mL)					
		Models: 1a8			Models: 1a5, 1a6 and 1a7.		
		Abundance Baetidae	Elminthidae	Fecal Coliforms	Abundance Baetidae	Scirtidae	Fecal Coliforms
3	5	3	2.00	7.0E+01	3	0	7.0E+01
4	5	13	2.00	4.5E+00	13	0	4.5E+00
13	5	37	0.00	3.3E+02	37	0	3.3E+02
15	3	0	3.00	3.3E+02	0	132	3.3E+02
19	5	0	0.00	7.9E+01	0	1	7.9E+01
20	4	2	0.00	4.9E+01	2	1	4.9E+01
24	3	43	16.00	2.4E+02	43	0	2.4E+02
25	5	16	2.00	4.9E+01	16	0	4.9E+01
26	6	2	0.00	4.9E+01	2	2	4.9E+01
27	3	38	3.00	4.6E+02	38	3	4.6E+02
28	6	3	1.00	4.9E+01	3	1	4.9E+01
29	6	2	1.00	3.3E+01	2	0	3.3E+01
30	3	1	5.00	7.9E+01	1	0	7.9E+01
31	6	1	1.00	3.3E+01	1	1	3.3E+01
32	3	0	0.00	4.9E+01	0	0	4.9E+01
33	6	3	2.00	3.3E+01	3	1	3.3E+01
34	2	53	4.00	1.7E+05	53	1	1.7E+05
35	2	7	0.00	1.6E+04	7	4	1.6E+04
36	2	189	1.00	5.4E+05	189	4	5.4E+05
37	2	45	0.00	6.0E+03	45	0	6.0E+03
38	2	108	0.00	1.1E+05	108	2	1.1E+05
39	1	151	0.00	1.7E+04	151	0	1.7E+04
40	1	31	0.00	1.7E+05	31	0	1.7E+05
41	1	17	0.00	1.1E+05	17	0	1.1E+05
42	1	19	0.00	4.6E+04	19	0	4.6E+04
43	5	1	1.00	1.3E+02	1	0	1.3E+02
44	5	3	0.00	7.0E+01	3	0	7.0E+01
45	5	4	0.00	4.6E+02	4	0	4.6E+02
46	3	30	4.00	1.7E+01	30	0	1.7E+01
47	6	0	1.00	2.3E+01	0	0	2.3E+01
49	5	140	1.00	1.7E+01	140	0	1.7E+01
50	3	0	0.00	3.3E+01	0	0	3.3E+01
51	5	0	0.00	4.5E+00	0	0	4.5E+00

Symbology:

Land use: 1 Urban area; 2 Suburban area, pastures and crops; 3 Pastures; 4 Bare soil; 5 Native vegetation; 6 Forest vegetation

Fulfilment of regulation

Non-fulfilment of regulation

Place where the first rule (branch) of the decision tree is applicable

Place where the second rule(branch) of the decision tree is applicable

Table D5. Verification of the fulfilment of the agriculture and livestock water use regulations associated with fecal coliforms according the decision tree models (DTMs), before the optimization process.

Sampled Points	Land Use	Agriculture fecal coliform regulation (threshold = 1.00E+03 MPN/100 mL)						
		Model: 2ap3				Models: 2a1, 2a2, 2a3, 2a4, 2a5, 2a6.		
		Presence/Absence			Fecal	Abundance		
		Perlidae	Baetidae	Leptophlebiidae	Coliforms	Baetidae	Perlidae	Fecal Coliforms
3	5	-	P	P	7.0E+01	3	0	7.0E+01
4	5	P	P	-	4.5E+00	13	1	4.5E+00
13	5	P	P	P	3.3E+02	37	7	3.3E+02
15	3	-	-	-	3.3E+02	0	0	3.3E+02
19	5	-	-	-	7.9E+01	0	0	7.9E+01
20	4	-	P	-	4.9E+01	2	0	4.9E+01
24	3	P	P	-	2.4E+02	43	6	2.4E+02
25	5	P	P	P	4.9E+01	16	1	4.9E+01
26	6	-	P	P	4.9E+01	2	0	4.9E+01
27	3	P	P	P	4.6E+02	38	4	4.6E+02
28	6	P	P	P	4.9E+01	3	1	4.9E+01
29	6	P	P	-	3.3E+01	2	1	3.3E+01
30	3	P	P	-	7.9E+01	1	1	7.9E+01
31	6	-	P	-	3.3E+01	1	0	3.3E+01
32	3	-	-	-	4.9E+01	0	0	4.9E+01
33	6	P	P	-	3.3E+01	3	1	3.3E+01
34	2	-	P	-	1.7E+05	53	0	1.7E+05
35	2	-	P	-	1.6E+04	7	0	1.6E+04
36	2	-	P	-	5.4E+05	189	0	5.4E+05
37	2	-	P	-	6.0E+03	45	0	6.0E+03
38	2	-	P	-	1.1E+05	108	0	1.1E+05
39	1	-	P	-	1.7E+04	151	0	1.7E+04
40	1	-	P	-	1.7E+05	31	0	1.7E+05
41	1	-	P	-	1.1E+05	17	0	1.1E+05
42	1	-	P	-	4.6E+04	19	0	4.6E+04
43	5	P	P	-	1.3E+02	1	1	1.3E+02
44	5	-	P	P	7.0E+01	3	0	7.0E+01
45	5	-	P	-	4.6E+02	4	0	4.6E+02
46	3	P	P	P	1.7E+01	30	1	1.7E+01
47	6	-	-	-	2.3E+01	0	0	2.3E+01
49	5	P	P	P	1.7E+01	140	9	1.7E+01
50	3	-	-	-	3.3E+01	0	0	3.3E+01
51	5	-	-	-	4.5E+00	0	0	4.5E+00

Symbology:

Land use: 1 Urban area; 2 Suburban area, pastures and crops; 3 Pastures; 4 Bare soil; 5 Native vegetation; 6 Forest vegetation

Fulfilment of regulation (blue)
Non-fulfilment of regulation (orange)
Place where the first rule (branch) of the decision tree is applicable (green)
Place where the second rule(branch) of the decision tree is applicable (light green)

Table D6. Verification of the fulfilment of the recreational with primary contact Ecuadorian water use regulations associated with fecal coliforms according the decision tree models (DTMs), with new dataset taken in July of 2015 (dry season) and March of 2016 (rainy season).

Sampled Points	Season	Land Use	Recreational fecal coliform regulation (threshold = 2.00E+02 MPN/100 mL)					
			Models: 1a8			Models: 1a5, 1a6 and 1a7.		
			Abundance Baetidae	Elminthidae	Fecal Coliforms	Abundance Baetidae	Scirtidae	Fecal Coliforms
01V	Rainy	5	4	83	6.8E+00	4	0	6.8E+00
03V	Rainy	6	1	13	4.5E+00	1	0	4.5E+00
05V	Rainy	6	2	4	4.5E+00	2	0	4.5E+00
06V	Rainy	6	16	0	4.0E+01	16	0	4.0E+01
07V	Rainy	3	5	3	2.3E+02	5	0	2.3E+02
09V	Rainy	2	10	0	1.7E+04	10	0	1.7E+04
11V	Rainy	1	48	4	2.7E+04	48	0	2.7E+04
13V	Rainy	1	38	2	9.3E+04	38	0	9.3E+04
15V	Rainy	5	2	0	1.8E+00	2	0	1.8E+00
16V	Rainy	5	0	37	1.8E+00	0	0	1.8E+00
17V	Rainy	5	0	25	4.5E+00	0	0	4.5E+00
01V	Dry	5	2	16	2.0E+00	2	1	2.0E+00
03V	Dry	6	0	2	4.0E+00	0	0	4.0E+00
05V	Dry	6	3	0	1.8E+00	3	0	1.8E+00
06V	Dry	6	7	8	1.3E+01	7	0	1.3E+01
07V	Dry	3	58	1	1.7E+02	58	0	1.7E+02
08V	Dry	2	67	2	5.4E+03	67	0	5.4E+03
09V	Dry	2	376	1	3.3E+03	376	0	3.3E+03
10V	Dry	2	463	2	2.2E+04	463	5	2.2E+04
11V	Dry	1	91	1	1.7E+05	91	0	1.7E+05
12V	Dry	1	121	0	1.1E+06	121	0	1.1E+06
13V	Dry	1	8	0	7.0E+05	8	0	7.0E+05
15V	Dry	5	0	0	1.8E+00	0	0	1.8E+00
16V	Dry	5	2	17	1.8E+00	2	0	1.8E+00
17V	Dry	5	1	11	2.0E+00	1	2	2.0E+00

Symbology:

Land use: 1 Urban area; 2 Suburban area, pastures and crops; 3 Pastures; 4 Bare soil; 5 Native vegetation; 6 Forest vegetation

Fulfilment of regulation

Non-fulfilment of regulation

Place where the first rule (branch) of the decision tree is applicable

Place where the second rule(branch) of the decision tree is applicable

Table D7. Verification of the fulfilment of the agriculture and livestock Ecuadorian water use regulations associated with fecal coliforms according the decision tree models (DTMs), with new dataset taken in July of 2015 and March of 2016.

Sampled Points	Season	Land Use	Agriculture fecal coliform regulation (threshold = 1.00E+03 MPN/100 mL)						
			Model: 2ap3				Models: 2a1, 2a2, 2a3, 2a4, 2a5, 2a6.		
			Presence/Absence			Fecal Coliforms	Abundance		Fecal Coliforms
			Perlidae	Baetidae	Leptophlebiidae		Baetidae	Perlidae	
01V	Rainy	5	-	P	P	6.8E+00	4	-	6.8E+00
03V	Rainy	6	-	P	-	4.5E+00	1	-	4.5E+00
05V	Rainy	6	-	P	-	4.5E+00	2	-	4.5E+00
06V	Rainy	6	-	P	-	4.0E+01	16	-	4.0E+01
07V	Rainy	3	-	P	-	2.3E+02	5	-	2.3E+02
09V	Rainy	2	-	P	-	1.7E+04	10	-	1.7E+04
11V	Rainy	1	-	P	-	2.7E+04	48	-	2.7E+04
13V	Rainy	1	-	P	-	9.3E+04	38	-	9.3E+04
15V	Rainy	5	-	P	-	1.8E+00	2	-	1.8E+00
16V	Rainy	5	-	-	P	1.8E+00	0	-	1.8E+00
17V	Rainy	5	-	-	P	4.5E+00	0	-	4.5E+00
01V	Dry	5	-	P	P	4.0E+00	2	-	4.0E+00
03V	Dry	6	-	-	-	1.8E+00	0	-	1.8E+00
05V	Dry	6	-	P	-	1.3E+01	3	-	1.3E+01
06V	Dry	6	-	P	-	1.7E+02	7	-	1.7E+02
07V	Dry	3	-	P	-	5.4E+03	58	-	5.4E+03
08V	Dry	2	-	P	-	3.3E+03	67	-	3.3E+03
09V	Dry	2	-	P	-	2.2E+04	376	-	2.2E+04
10V	Dry	2	-	P	-	1.7E+05	463	-	1.7E+05
11V	Dry	1	-	P	-	1.1E+06	91	-	1.1E+06
12V	Dry	1	-	P	-	7.0E+05	121	-	7.0E+05
13V	Dry	1	-	P	-	1.8E+00	8	-	1.8E+00
15V	Dry	5	-	-	-	1.8E+00	0	-	1.8E+00
16V	Dry	5	-	P	P	2.0E+00	2	-	2.0E+00
17V	Dry	5	-	P	P	0.0E+00	1	-	0.0E+00

Symbology:

Land use: 1 Urban area; 2 Suburban area, pastures and crops; 3 Pastures; 4 Bare soil; 5 Native vegetation; 6 Forest vegetation

Fulfilment of regulation

Non-fulfilment of regulation

Place where the first rule (branch) of the decision tree is applicable

Place where the second rule(branch) of the decision tree is applicable

Table D8. Calculation of the variation of correctly classified instances (CCI) and Kappa statistics in the recreational fecal regulation models, in which three cross validations were manually applied

Dataset	Cross validation		Cost Matrix Weights (CMW)				CCI [e] (%)			Kappa statistics		
	Folds	Pruning confidence factor (PCF)	TP [a]	FN [b]	FP [c]	TN [d]	CCI	±	sd	Kappa	±	sd
Abundance	3	0.25					84.8	±	10.5	0.7	±	0.2
Abundance	3	0.10					81.8	±	9.1	0.6	±	0.2
Abundance	3	0.25	0	1	2	0	78.8	±	18.9	0.6	±	0.4
Abundance	3	0.10	0	1	2	0	78.8	±	18.9	0.6	±	0.4
Abundance	3	0.25	0	1	3	0	69.7	±	22.9	0.4	±	0.4
Abundance	3	0.10	0	1	3	0	75.8	±	13.9	0.5	±	0.3
Abundance	3	0.25	0	1	5	0	69.7	±	22.9	0.4	±	0.4
Abundance	3	0.10	0	1	5	0	69.7	±	22.9	0.4	±	0.4
Abundance	3	0.25	0	1	7	0	54.5	±	15.7	0.1	±	0.3
Abundance	3	0.10	0	1	7	0	54.5	±	15.7	0.2	±	0.3
Presence-absence	3	0.25					72.7	±	9.1	0.4	±	0.2
Presence-absence	3	0.10					63.6	±	9.1	0.3	±	0.2
Presence-absence	3	0.25	0	1	2	0	54.5	±	9.1	0.1	±	0.2
Presence-absence	3	0.25	0	1	3	0	54.5	±	9.1	0.1	±	0.2
Presence-absence	3	0.25	0	1	5	0	51.5	±	10.5	0.1	±	0.2
Presence-absence	3	0.25	0	1	7	0	45.5	±	0.0	-0.0	±	0.0
Presence-absence	3	0.25	0	1	10	0	45.5	±	0.0	0.0	±	0.0

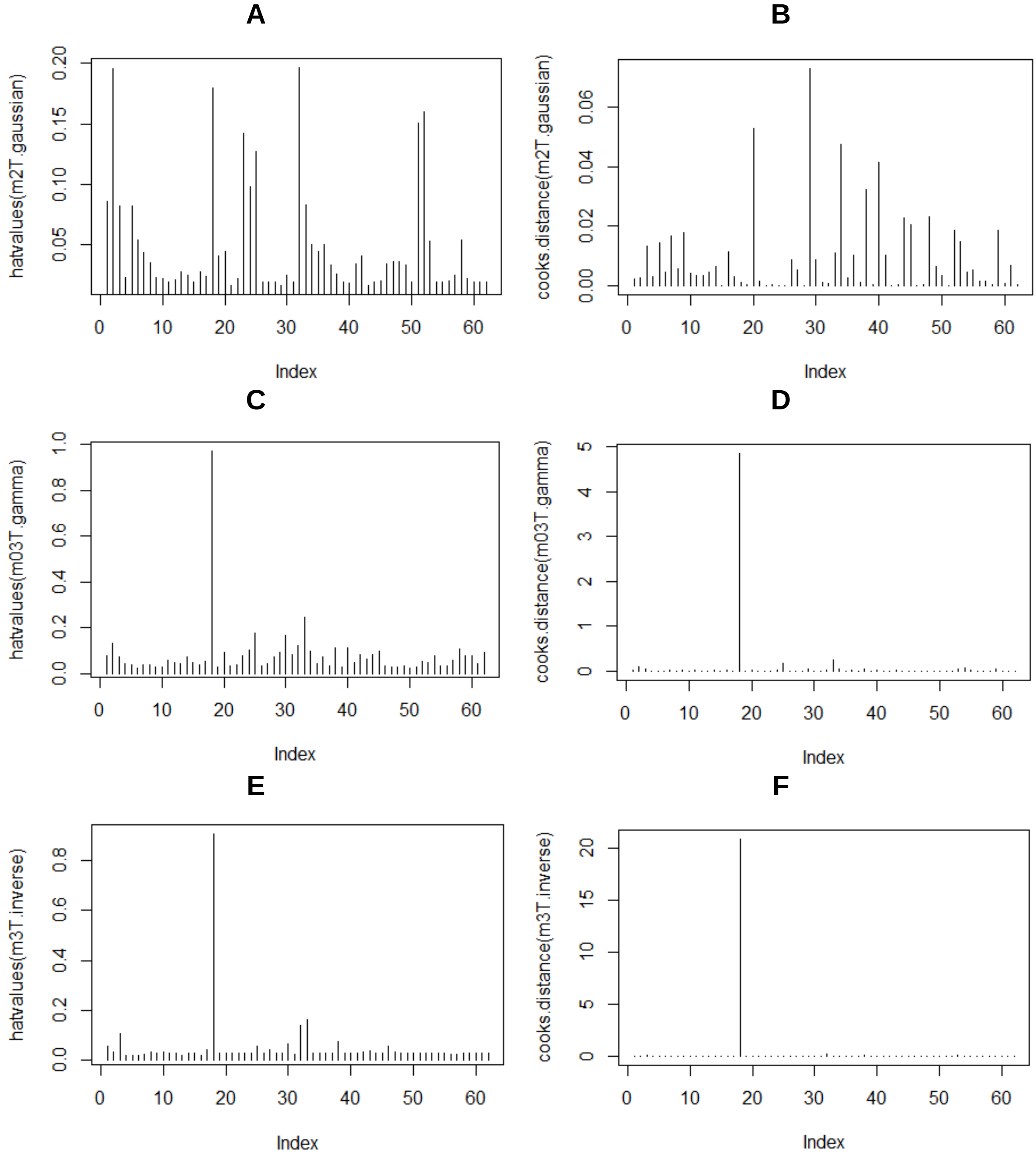
[a] TP =True positives
[b] FN =False negative
[c] FP =False positive
[d] TN =True negative
[e] CCI= Correctly classified instances

Table D9. Calculation of the variation of correctly classified instances (CCI) and Kappa statistics in the agriculture fecal regulation models, in which three cross validations were manually applied

Dataset	Cross validation		Cost Matrix Weights (CMW)				CCI [e] (%)			Kappa statistics		
	Folds	Pruning confidence factor (PCF)	TP [a]	FN [b]	FP [c]	TN [d]	CCI	±	sd	Kappa	±	sd
Abundance	3	0.25					78.8	±	13.9	0.5	±	0.3
Abundance	3	0.10					78.8	±	13.9	0.5	±	0.3
Abundance	3	0.25	0	1	2	0	84.8	±	18.9	0.3	±	0.4
Abundance	3	0.25	0	1	3	0	84.8	±	18.9	0.3	±	0.4
Abundance	3	0.25	0	1	5	0	84.8	±	18.9	0.3	±	0.4
Abundance	3	0.25	0	1	7	0	84.8	±	18.9	0.3	±	0.4
Abundance	3	0.25	0	1	10	0	84.8	±	18.9	0.3	±	0.4
Presence-absence	3	0.25					75.8	±	5.2	0.4	±	0.2
Presence-absence	3	0.10					72.7	±	9.1	0.2	±	0.2
Presence-absence	3	0.25	0	1	2	0	72.7	±	9.1	0.4	±	0.3
Presence-absence	3	0.25	0	1	3	0	81.8	±	9.1	0.6	±	0.2
Presence-absence	3	0.25	0	1	5	0	57.6	±	5.2	0.2	±	0.2
Presence-absence	3	0.25	0	1	7	0	57.6	±	5.2	0.2	±	0.2
Presence-absence	3	0.25	0	1	10	0	57.6	±	5.2	0.2	±	0.2

[a] TP =True positives
[b] FN =False negative
[c] FP =False positive
[d] TN =True negative
[e] CCI= Correctly classified instances

Appendix E – Supporting information for chapter 6



Index = influcen point
Index 18 corresponds to site Ta02

Fig. E1. Plots of influential points with the complete dataset, using Cook's distance and hat elements with different kind of GLM: (**A - B**) Gaussian, (**C - D**) Gamma, and (**E - F**) Inverse Gaussian.

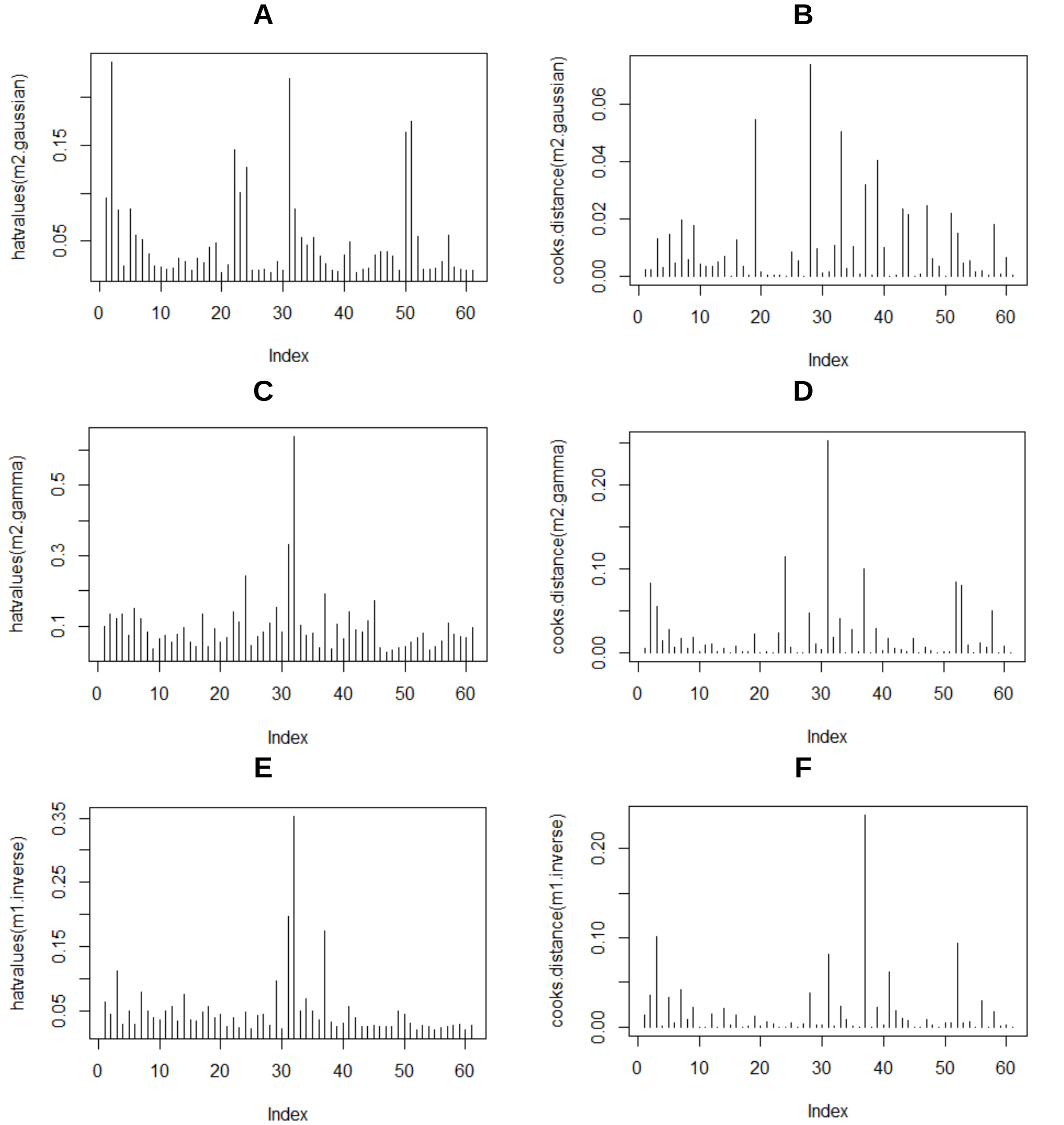
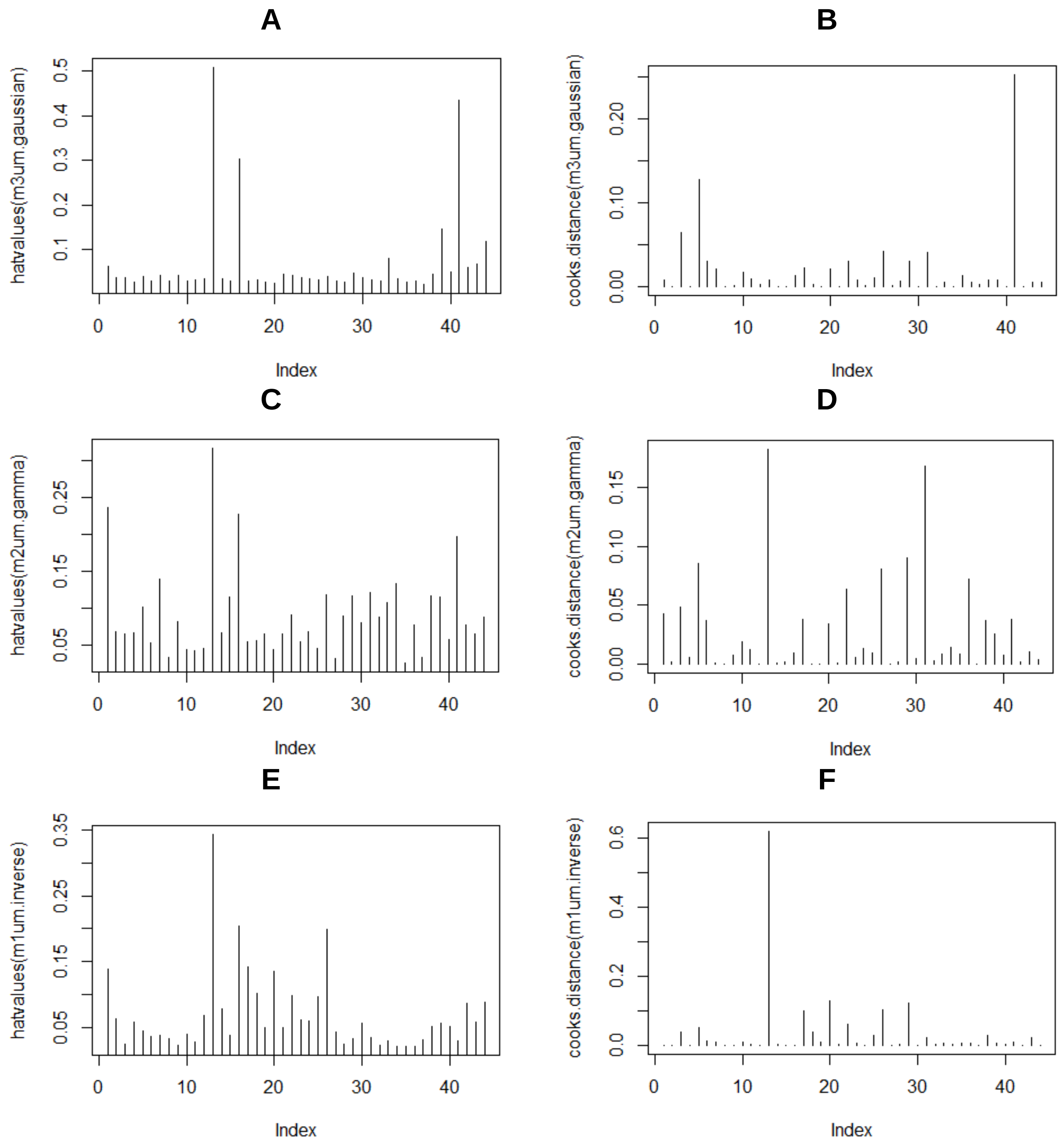


Fig. E2. Plots of influential points with the without-outliers (Ta02) dataset, using Cook's distance and hat elements with different kind of GLM: (**A - B**) Gaussian, (**C - D**) Gamma, and (**E - F**) Inverse Gaussian.



Index = influcen point
Index 13 corresponds to site Ta01 during rainy season.

Fig. E3. 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.

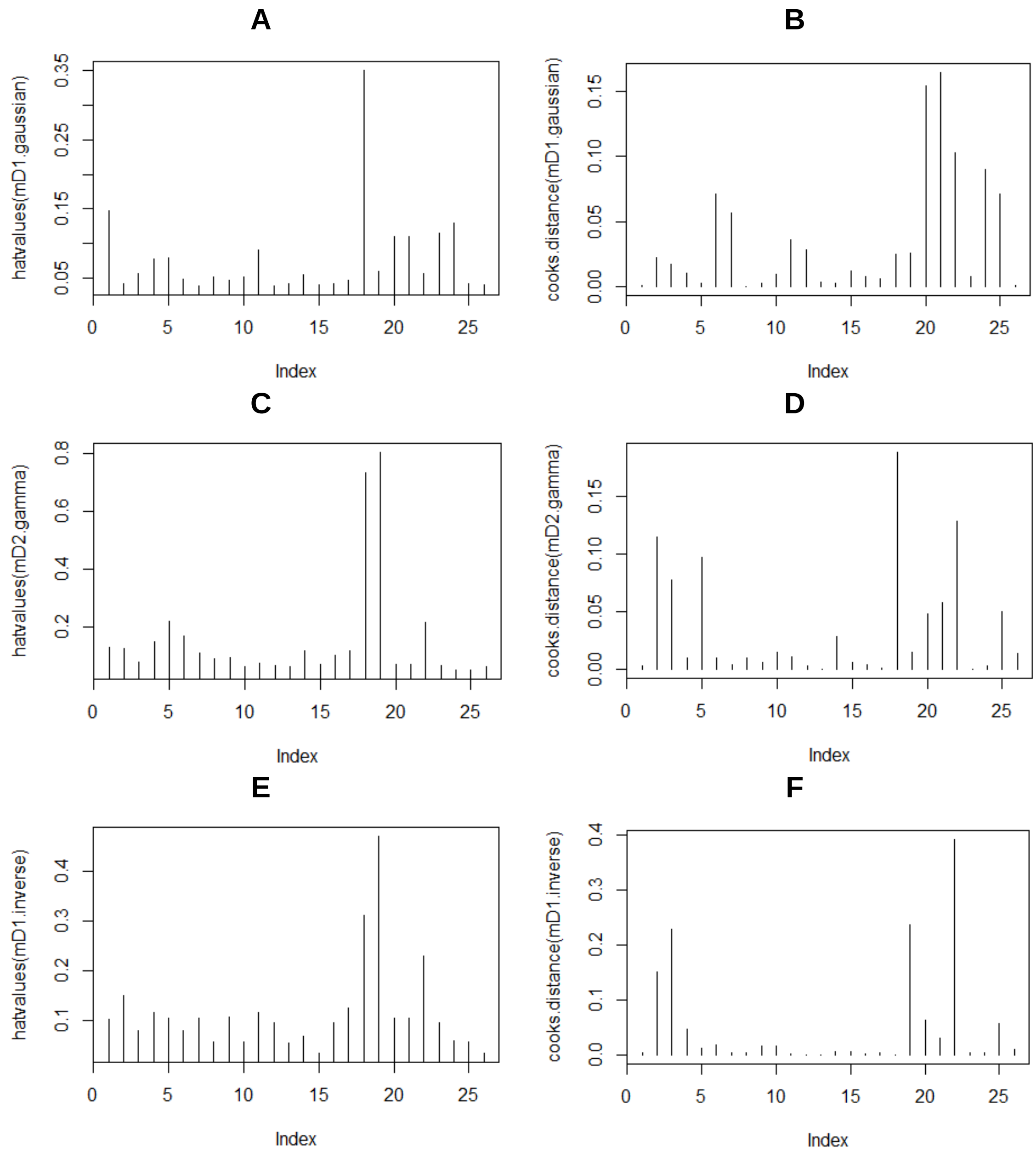


Fig. E4. Plots of influential points with the dry-season dataset, using Cook's distance and hat elements with different kind of GLM: (**A - B**) Gaussian, (**C - D**) Gamma, and (**E - F**) Inverse Gaussian.

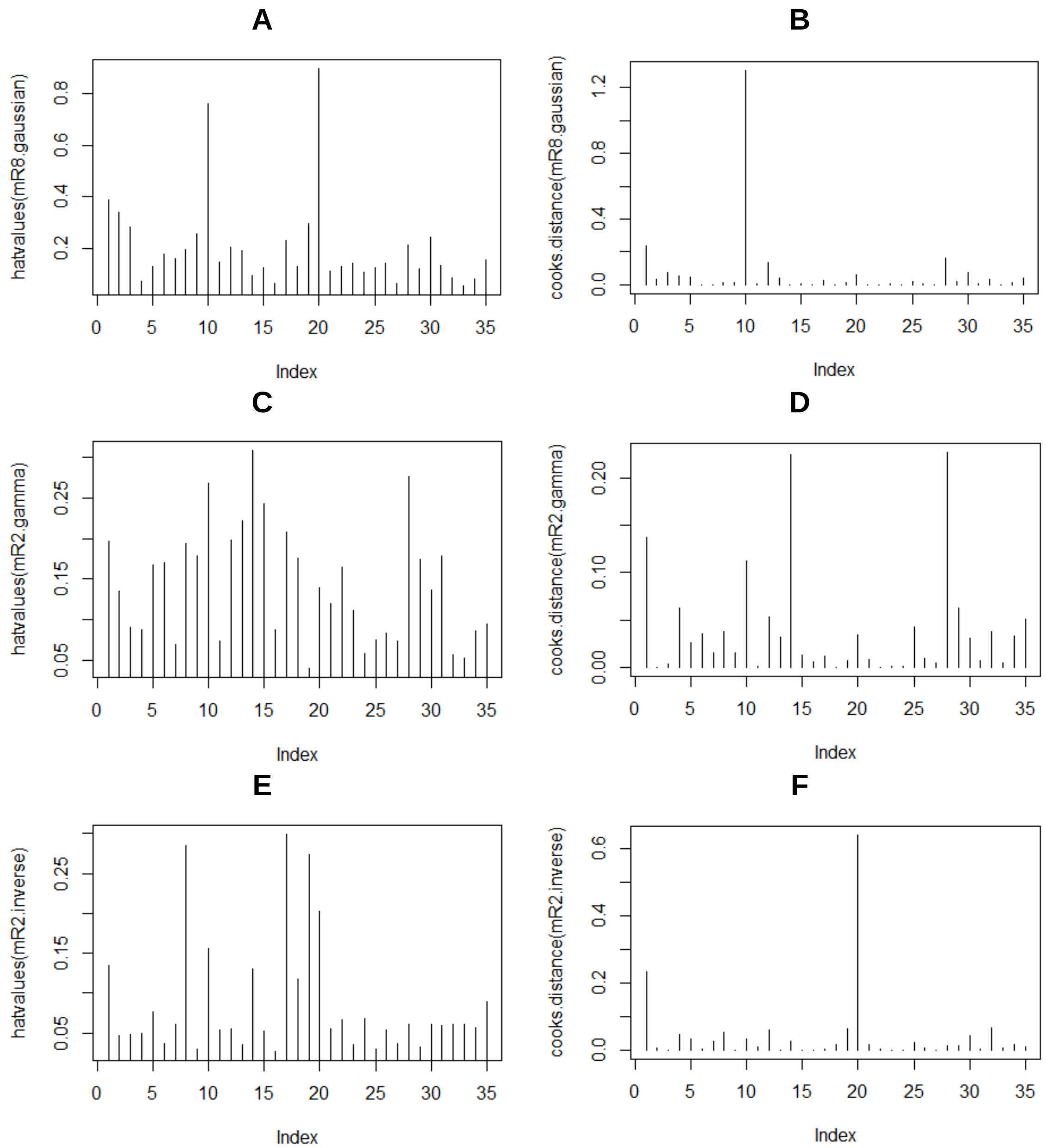


Fig. E5. Plots of influential points with the rainy-season dataset, using Cook's distance and hat elements with different kind of GLM: (**A - B**) Gaussian, (**C - D**) Gamma, and (**E - F**) **I**nverse Gaussian

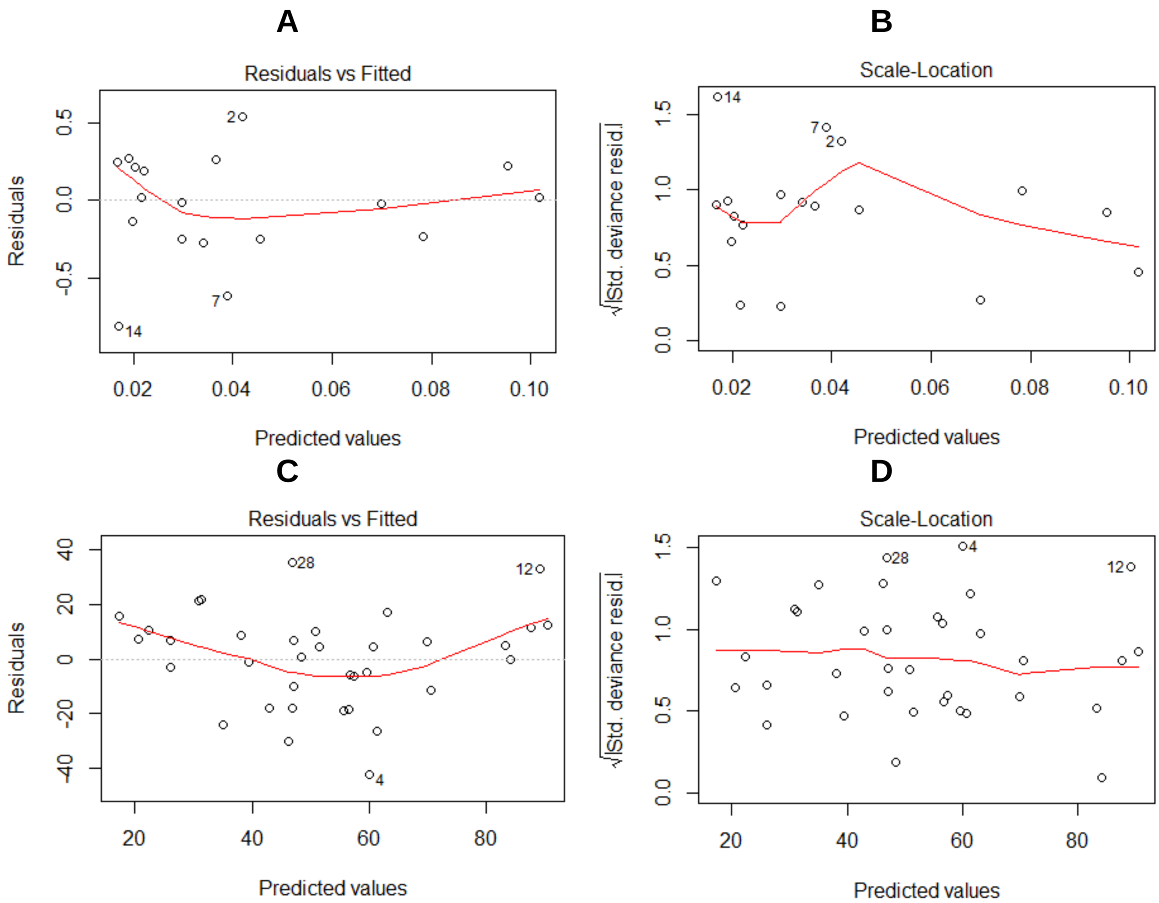


Fig. E6. Plots of Residual vs. Fitted of the best models: (**A**) Dry season, (**B**) Rainy season, and plots of Scale-Location of the best models: (**C**) Dry season, (**D**) Rainy season.

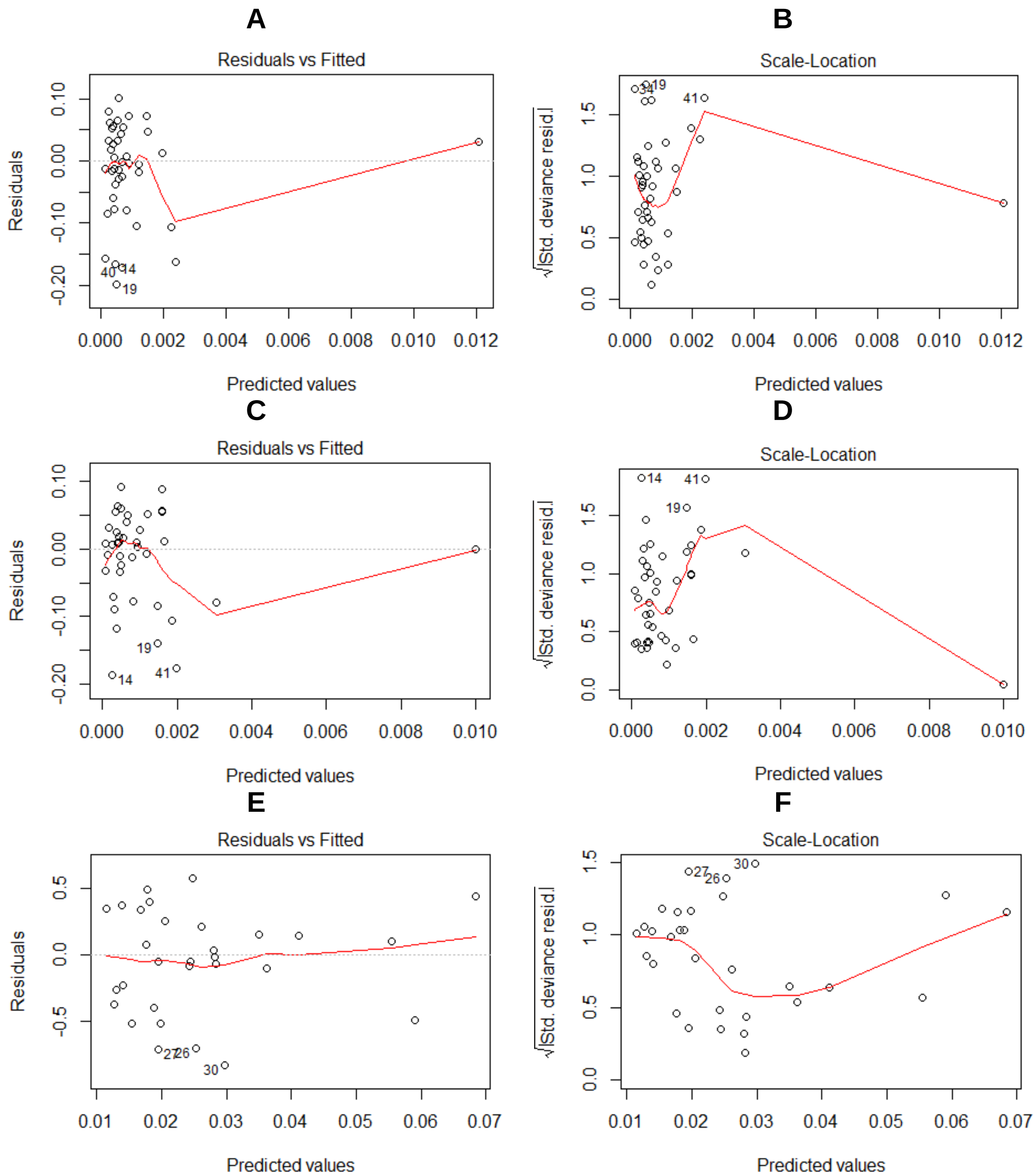


Fig. E7. Plots of Residual vs. Fitted of the best models: (**A**) with the complete dataset, (**B**) with the without-outliers dataset, (**C**) with the main-rivers dataset, and plots of Scale-Location of the best models: (**D**) with the complete dataset, (**E**) the without-outliers dataset, and (**F**) with the main-rivers dataset.

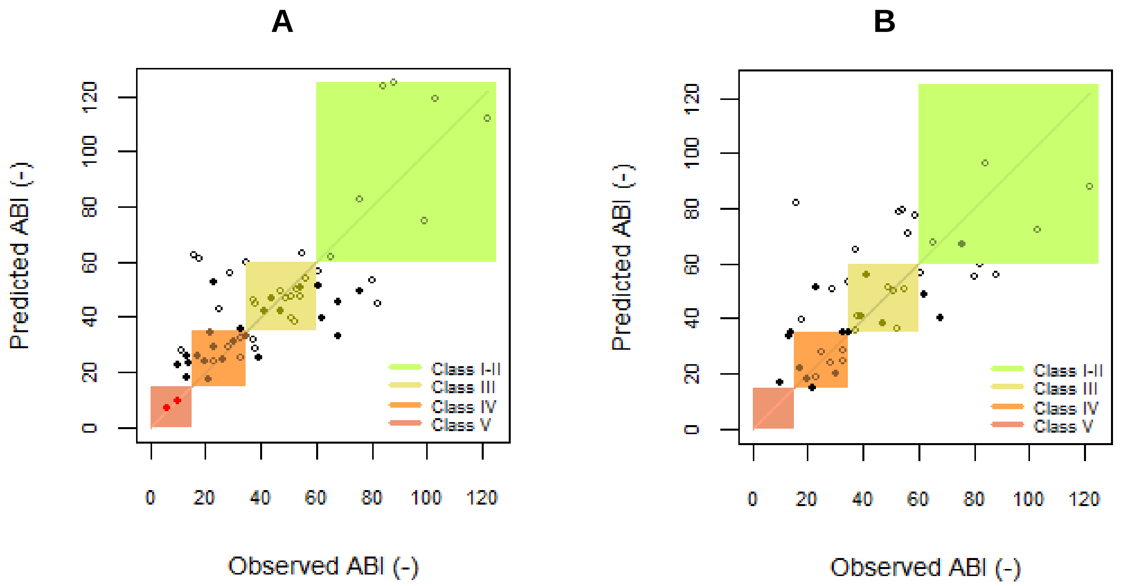


Fig. E8. Graph comparing simulated index and field data for models developed from: (**A**) the without-outlier dataset, and (**B**) the main-rivers dataset. The filled dots and the empty dots are the sampling sites taken during dry and rainy seasons respectively.

Table E1. Results of the variables difference between seasons

Parameter	Normality tests: Levene Test	Normality tests: Shapiro Test	Paired Student test: Two sided	Paired Student test: Greater during rainy season	Paired Student test: Greater during dry season	Paired Wilcoxon test: Two sided	Paired Wilcoxon test: Greater during rainy season	Paired Wilcoxon test: Greater during 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/100mL)	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 E2. Best model outcomes: dry season

Explanatory Variables	Regression Parameters	Models obtained from the dry-season dataset (without outliers)							
		Complete data		Fold-1		Fold-2		Fold-3	
		Inverse Gaussian model: mD1.inverse		Inverse Gaussian model: mD9.3fcv1a2.inverse		Inverse Gaussian model: mD6.3fcv1a3.gamma		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 E3. Best model outcomes: rainy season

Explanatory Variables	Regression Parameters	Models obtained from the rainy-season dataset (without outliers)							
		Complete data		Fold-1		Fold-2		Fold-3	
		Gaussian model mR8.gaussian		Gaussian model mr29.3fcv1a2. gaussian		Gaussian model mR07.3fcv1a3. gaussian		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 E4. Best model outcomes with the complete dataset

Explanatory Variables	Regression Parameters	Models obtained from both season with the complete dataset							
		Complete data		Fold-1		Fold-2		Fold-3	
		Gamma model: m03T.gamma		Gamma model m2.3fcv1a2T.gamma		Inverse Gaussian model m5T.3fcv1a3T.inverse		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 E5. Best model outcomes without-outliers dataset

Explanatory Variables	Regression Parameters	Models obtained from both season without-outliers dataset							
		Complete data		Fold-1		Fold-2		Fold-3	
		Gamma model m2.gamma		Gamma model m3.3fcv1a2.gamma		Inverse Gaussian model m2.3fcv1a3.inverse		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 E6. Best model outcomes with the main-rivers dataset

Explanatory Variables	Regression Parameters	Models obtained from both season with the main-rivers dataset							
		Complete data		Fold-1		Fold-2		Fold-3	
		Gamma model: m6um.gamma		Gaussian model: m2um3fcv1a2.gaussian		Gamma model: m13um3fcv1a3.gamma		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	

Appendix F – Supporting information for chapter 7

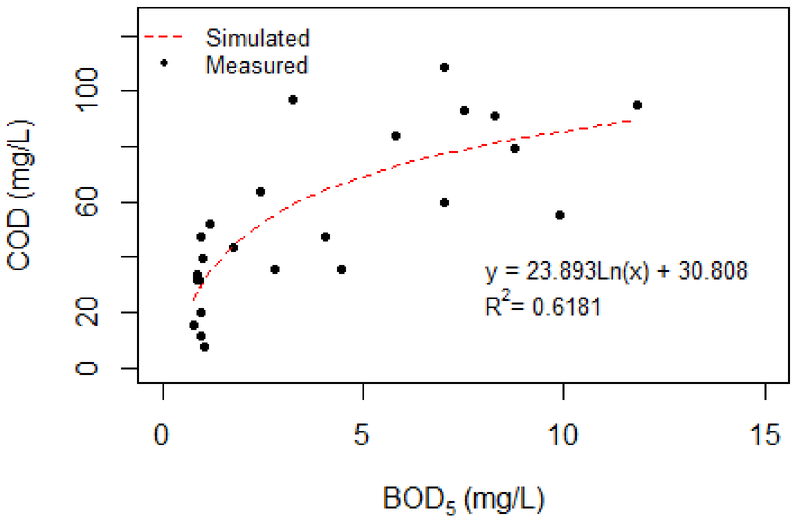


Fig. F1. Adjustment of the relationship between COD and BOD_5 to a logarithmic regression

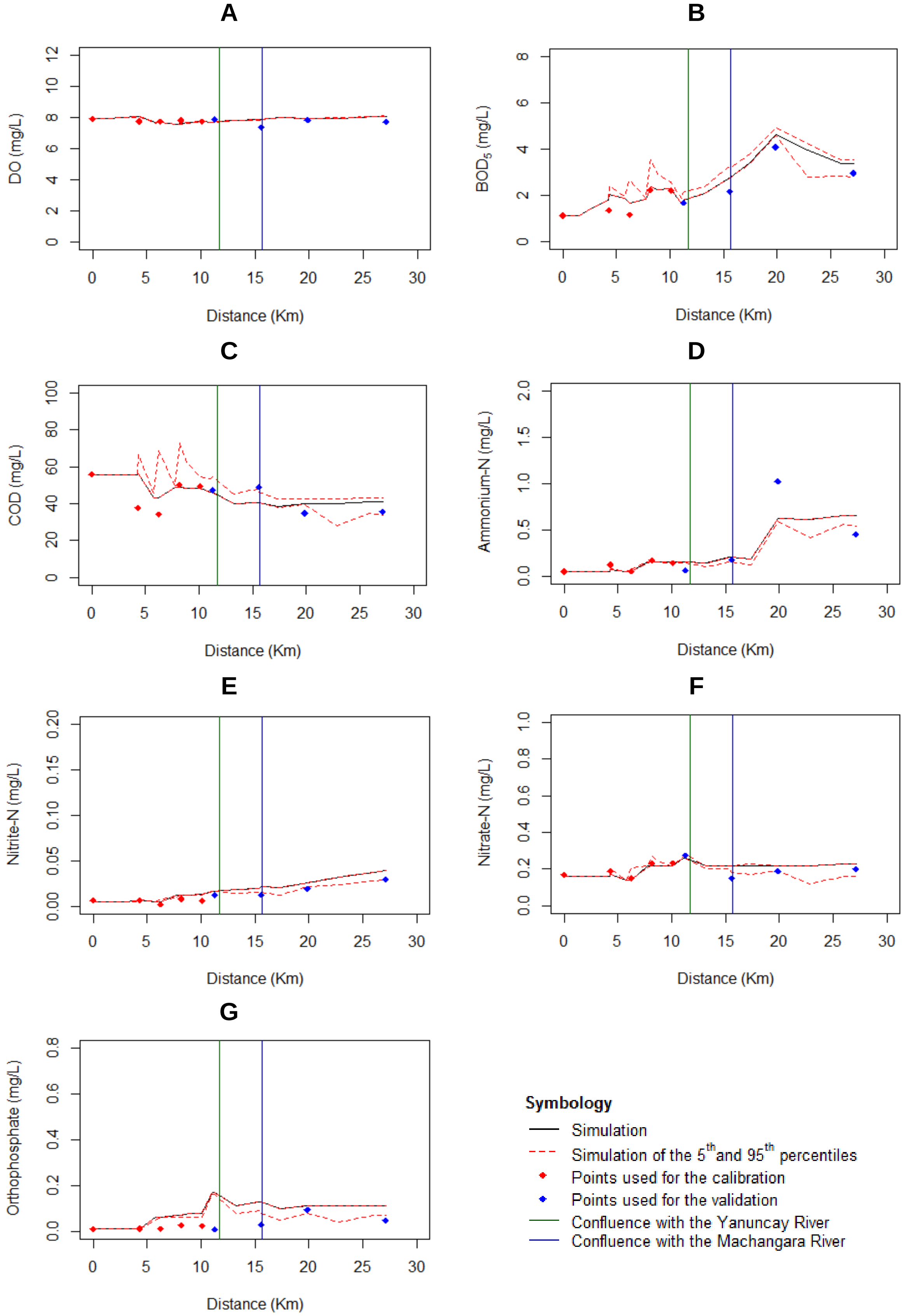


Fig. F2. Calibrated water quality model in the Tomebamba River during the rainy season for: (**A**) DO, (**B**) BOD_5, (**C**) COD, (**D**) Ammonium, (**E**) Nitrite, (**F**) Nitrate, (**G**) Orthophosphate.

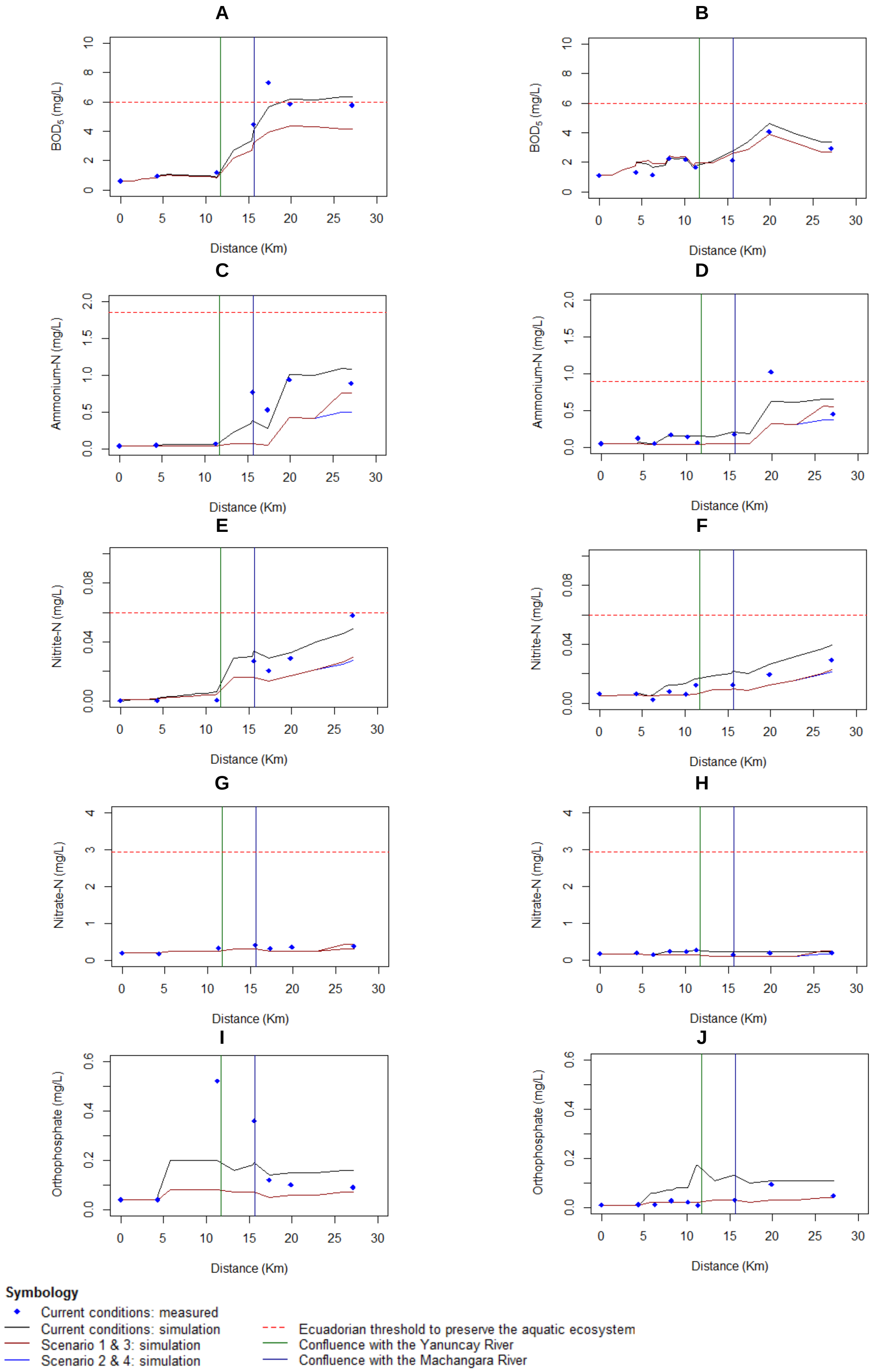


Fig. F3. Simulated concentrations along the Tomebemba and Cuenca Rivers. For dry season - Scenario 1 and 2: (**A**) BOD_5, (**C**) ammonium, (**E**) nitrite, (**G**) nitrate and (**I**) orthophosphate. For rainy season - Scenario 3 and 4: (**B**) BOD_5, (**D**) ammonium, (**F**) nitrite, (**H**) nitrate and (**J**) orthophosphate.

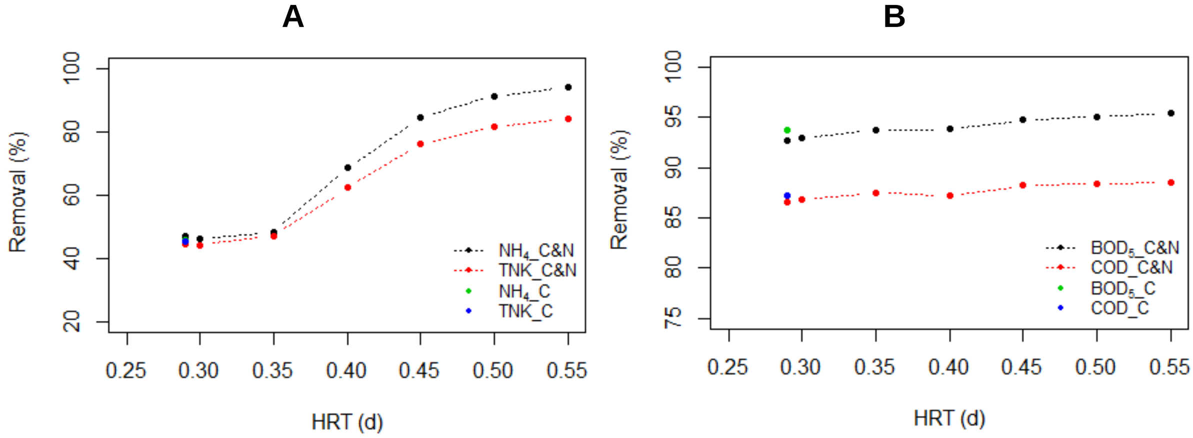


Fig. F4. Analysis of removal for the new G-WWTP according of hydraulic retention time (HRT) with the technologies of carbon (C) and carbon and nitrogen elimination (C&N) for: (A) nitrogen family, and (B) carbon family.

Table F1. Fraction of the COD used in the parametrization of the river water quality model (RWQM).

Measured variables		RWQM1 variable	Units	Fraction	Value
COD Soluble	Readily biodegradable soluble COD	SS	$gCOD.m-^{3}$	0.6	
	Fraction soluble of COD that is inert	SI	$gCOD.m^{-3}$	0.4	
COD particulate	Fraction particulate of COD that is organic material	XS[a]	$gCOD.m^{-3}$	0.65	
	Fraction particulate of COD that is inert	XI[a]	$gCOD.m^{-3}$	~0.35	
	Chlorophyll-a	XALG	$gCOD.m^{-3}$	0.4167	
	Heterotrophic biomass	XH	$gCOD.m^{-3}$		2
	First stage nitrifying bacteria	XN1	$gCOD.m^{-3}$		0.4
	Second stage nitrifying bacteria	XN2	$gCOD.m^{-3}$		0.2

[a] $X_I + X_S = COD_{Total} - COD_{soluble} - X_{ALG} - X_H - X_{N1} - X_{N2}$

Table F2. Range of the parameters considered in the calibration of the river water quality model.

Parameter		Units	Default	Adjusted value
Dry season:				
Base value for kla	kla_{base}	d^{-1}	1.0	0.10
Rainy season:				
Base value for kla	kla_{base}	d^{-1}	1.0	1.55

Table F3. Water quality model evaluation with different indices: R^2, RMSE and χ^2.

Variable	Dry season						Rainy season					
	Calibration			Validation			Calibration			Validation		
	R^2	RMSE	χ^2	R^2	RMSE	χ^2	R^2	RMSE	χ^2	R^2	RMSE	χ^2
Water depth	0.78	0.15	0.95	0.87	0.13	0.90	0.78	0.15	0.95	0.87	0.13	0.90
Flow velocity	0.86	0.26	0.86	0.74	0.11	0.90	0.86	0.26	0.86	0.74	0.11	0.90
DO	0.41	0.02	0.99	0.97	0.05	0.96	0.18	0.03	0.99	0.05	0.04	0.97
BOD_5	0.99	0.15	0.95	0.98	0.06	0.80	0.86	0.34	0.75	0.98	0.17	0.86
COD	0.99	0.09	0.54	0.75	0.02	0.54	0.01	0.27	0.01	0.24	0.14	0.21
Ammonium	0.99	0.31	0.90	0.96	0.14	0.92	0.76	0.24	0.99	0.66	0.71	0.84
Nitrite	0.99	0.09	0.93	0.48	0.42	0.01	0.67	0.58	0.09	0.01	0.32	0.01
Nitrate	0.98	0.45	0.89	0.75	0.63	0.72	0.98	0.04	0.99	0.84	0.39	0.95
Orthophosphates	0.93	0.50	0.87	0.83	0.48	0.98	0.40	2.82	0.79	0.67	9.17	0.19

[a] The completed dataset were applied for the calibration and validation of water depth and flow velocity. Namely, These variables were not calibrated and validated per season.