Promoters

Prof. Dr. ir Peter Goethals

Laboratory of Environmental Toxicology and Aquatic Ecology

Department of Applied Ecology and Environmental Biology

Dr. Pieter Boets

Laboratory of Environmental Toxicology and Aquatic Ecology

Department of Applied Ecology and Environmental Biology

Dean

Prof. dr. ir. G. Van Huylenbroeck

Rector

Prof. Dr. P. Van Cauwenberge

M.Sc. Javier Ernesto Holguin Gonzalez

Integrated ecological modelling for decision support in river management

Thesis submitted in fulfilment of the requirements for the degree of Doctor (PhD) in Applied Biological Sciences

Dutch translation of the title:

Geïntegreerde ecologische modellering voor beslissingsondersteunend rivierbeheer.

Please refer to this work as follows:

Holguin-Gonzalez, J.E., 2013. Integrated ecological modelling for decision support in river management. PhD thesis, Ghent University, Ghent, Belgium.

ISBN-number: 978-90-5989-640-6

This research was supported by an exchange funding from the European Union Alfa Project, the WATROPEC Project of the Flemish Government and the Special Research Fund of Ghent University (BOF01/WI0/611 and BOF09/24J/092) in Belgium.

Acknowledgements

First of all I would like to thank my promoters, Prof. Dr. ir. Peter Goethals and Dr. Pieter Boets at the Laboratory of Environmental Toxicology and Aquatic Ecology at Ghent University in Belgium for their support and guidance during this research.

Doing research is only possible when there are enough financial means. Therefore, I am very grateful to the representatives of the ALFA Programme "Latin-American – European Partnership For Water Education And Training (WETNET)", the Program "Master grants for candidates from developing countries", for their support of my studies in the Master program in Environmental Sanitation and my thesis research. Moreover, I like to thank the Research Unit of Aquatic Ecology that funded me via the WATROPEC Project of the Flemish Government, and the Special Research Fund of Ghent University (BOF01/WI0/611 and BOF09/24J/092) in Belgium.

In Colombia, I would like to thank the Environmental Technical Direction of the Environmental Authority in the Cauca Region (CVC), the Natural Resources and Environmental Engineering School of the Universidad del Valle (EIDENAR-Univalle), the Cauca River Modelling Project (CRMP) and Cinara Institute in Colombia, for the information, which was used to develop the research presented in this document. Special thanks are given to my friends and colleagues in the CRMP where I learned to love modelling.

In Belgium and Ecuador, I would like to thank the Flemish Interuniversity Council – University Development Cooperation VLIR-UOS and the Programs VLIR- IUC Cuenca and VLIR SRS, for financing this research in the framework of the Project South-South mobility allowances. Additionally, I would like to thank the Universidad de Cuenca and ETAPA in Ecuador for the opportunity to study their samples. I thank Prof. Gregory J. Pelletier for his valuable suggestions on how to use the software QUAL2Kw and the engineer Sandra Barros from ETAPA for her support in this study.

In Croatia, I would like to thank the Water Supply and Sanitation Company in the city of Varaždin in Croatia (VARKOM) for the opportunity to study their samples, the company ECO-TEHPROJECT in Croatia and the WATROPEC Project.

Finally, I would like to thank my wonderful parents for their love and their unconditional support during these long years, who always believed in and supported my choices in life. I cannot forget to say thanks to my friends and colleagues at the Laboratory of Aquatic Ecology in Ghent University. Moreover, my friends in Ghent from different nationalities, with whom I shared the huge pile of extra-ecological knowledge, in particular related to culture, gastronomy, religion, psychology, politics and of course dancing. Thank you guys for your company, good ideas and for your enormous support.

Table of contents

List of abbreviations

AIC: Akaike's Information Criterion

AICc: Second-order Akaike's Information Criterion

ANN: Artificial Neural Networks

ASM2d: Activated Sludge Model No. 2d

AUC: Area Under the Receiver-Operating-Characteristic Curve

BMWP: Biological Monitoring Working Party

BOD_5: Five-day Biological Oxygen Demand

CBOD: Carbonaceous Biological Oxygen Demand

CBODf: Fast Carbonaceous Biological Oxygen Demand

CBODs: Slow Carbonaceous Biological Oxygen Demand

CCI: Correctly Classified Instances

COD: Chemical Oxygen Demand

CPOM: Coarse Particulate Organic Matter

CRMP: Cauca River Modelling Project

CSM: Confidence Set of Models

CSTRS: Continuous Stirred Tank Reactor in Series

CT: Classification Trees

Cum. w_i: Cumulative Akaike weights

CVC: Environmental Authority in the Cauca Region

D: Water Depth

DHI: Danish Hydraulic Institute

dm: Modified Index of Agreement

DO saturation: Dissolved Oxygen Saturation

DO: Dissolved Oxygen

List of abbreviations (cont.)

DO-Prati: Dissolved Oxygen Prati index

EKBI: Expert Knowledge Based Index

EQO/EQS: Environmental Quality Objective/Standards

EQR: Ecological Quality Ratio

ETAPA: Water Supply and Sanitation Company in the city of Cuenca in Ecuador

EWQ: Ecological Water Quality

FC: Faecal Coliforms

FPOM: Fine Particulate Organic Matter

GAM: Generalized Additive Model

GLM: Generalized Linear Model

GLUE: Generalised Likelihood Uncertainty Estimation

HSI: Habitat Suitability Index

IBIAP: Biotic Integrity Index Using Aquatic Invertebrates

IEMF: Integrated Ecological Modelling Framework

IFAS: Integrated Fixed-film Activated Sludge

IFIM: Instream Flow Incremental Methodology

K: Cohen's Kappa Coefficient

LRM: Logistic Regression Model

MMIF: Multimetric Macroinvertebrate Index of Flanders

MPN: Most Probable Number

MSE: Mean Squared Error

MT: Model Trees

N: Total Nitrogen

List of abbreviations (cont.)

NBRM: Negative Binomial Regression Model

NH_4^+: Ammonia

NO_3: Nitrate

ORGN: Organic Nitrogen

ORGP: Organic Phosphorus

P/R: Production to Respiration Ratio

P: Total Phosphorous

PCA: Principal Component Analysis

PHABSIM: Physical Habitat Simulation Model

PO_4: Phosphate

PTS: Pollution Tolerance Scores

r: Pearson Correlation Coefficient

R^2: Determination Coefficient

RCC: River Continuum Concept

ROC curve: Receiver Operating Characteristics Curve

RT: Regression Trees

RWQM1: River Water Quality Model No.1

SD: Standard Deviation

SOD: Sediment Oxygen Demand

TSS: Total Suspended Solids

Univalle: Del Valle University

V: Water Velocity

VBA: Visual Basic for Applications

VARKOM: Water Supply and Sanitation Company in the City of Varaždin in Croatia

List of abbreviations (cont.)

WATROPEC: Water Treatment Optimization with Ecological Criteria

WFD: European Water Framework Directive

WFD-Explorer: Water Framework Directive Explorer

WPIs: Water Pollution Indices

WQ: Water Quality

WQIs: Water Quality Indices

WWTP: Wastewater Treatment Plant

Chapter 1: General introduction

1.1 Research context

Environmental managers are constantly driven by politics searching for an optimal balance between habitat conservation and economics. The evaluation of the impact of basin management plans and pollution control and sanitation programs on the river water quality is not straightforward. It is often unclear which combination of measures is most effective to reach this optimal balance. Therefore, the use of models to simulate physicochemical, hydromorphological and ecological river conditions is a key aspect in integrated water resources management. Thus, there is a need for the development of practical (modelling) tools to understand the elements that affect the ecological state of a river system and to predict how they will respond under different management policies. However, most traditional modelling frameworks are not able to meet these requirements as models tend to represent individual processes and to run independently (Kraft, 2011). Thus, model integration is required to perform comprehensive evaluations which would be impossible when analysing each individual component of the system separately.

In this PhD study, a newly developed conceptual framework for assessing ecological degradation in rivers and streams generated by physicochemical pollution and hydromorphological disturbances is presented. The proposed framework, called *Integrated Ecological Modelling Framework* (IEMF), considers the following conceptual elements: driving forces, pressures, physicochemical, hydromorphological and ecological state and response (Fig 1.1). The IEMF allows considering simultaneously the impact of different river pressures, such as the discharge of wastewaters and habitat degradation caused by changes in the hydromorphological conditions, on the ecological water quality. The IEMF has four basic modelling components: (1) a model characterising the processes of the WWTP; (2) a river water quantity model; (3) a physicochemical river water quality model and; (4) a river ecological model. This last component includes habitat suitability models for selected macroinvertebrate groups and ecological assessment models based on a macroinvertebrate biotic indices.

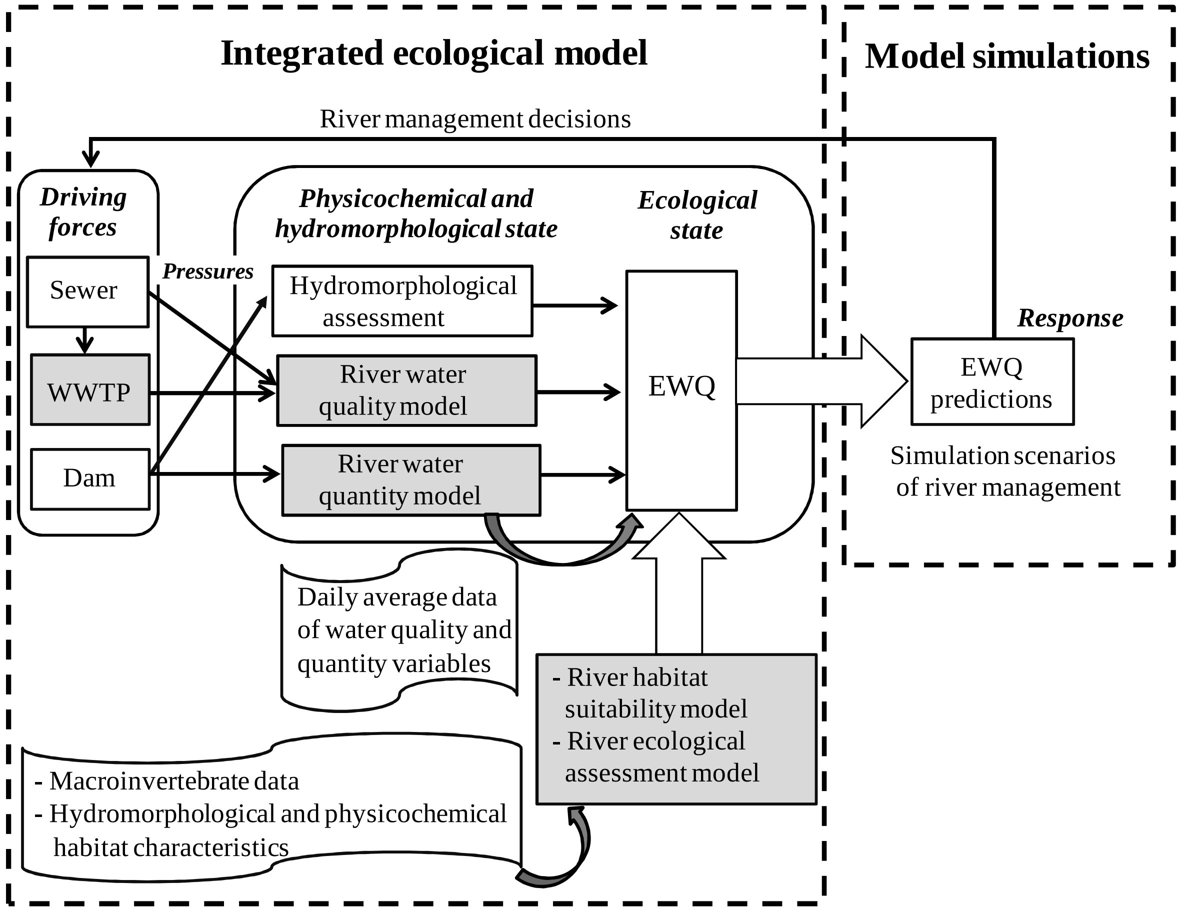


Fig. 1.1. Proposed framework developed in this PhD study for modelling river ecological water quality. This framework is called *Integrated Ecological Modelling Framework* (IEMF). The conceptual elements considered in the IEMF are presented in italics (i.e. driving forces, pressures, physicochemical, hydromorphological and ecological state and response). The four basic modelling components are found in grey boxes. (WWTP=wastewater treatment plant; EWQ=ecological water quality).

The physicochemical impacts on the river ecology considered in the IEMF are related with the discharge of treated (after wastewater treatment plants (WWTP)) or untreated wastewater (sewer discharge). Hydromorphological disturbances on river biota caused by dams or changes in the water course, current velocity, water depth, riverbed sediment composition and bank structure were considered in the IEMF. Regarding biological elements in the IEMF, this research focused on biological assessment based on a group of organisms collectively known as macroinvertebrates – organisms without a backbone, such as larval insects, crayfish, clams and snails. The biological assessment considered biotic indices derived from the occurrence and abundance of macroinvertebrate taxa and their sensitivity to organic pollution. Additionally, habitat suitability conditions for selected species of macroinvertebrates were evaluated.

The Driving force–Pressure–State-Impact-Response (DPSIR) (EEA, 1999) framework was selected as the basis for IEMF for the European Water Framework Directive (WFD) purposes, since many of the tasks required by the WFD refer directly to the elements of the DPSIR framework. The DPSIR framework was adopted by the European Environmental Agency and is based on the concept of causality chains for data synthesis, which links environmental information using indicators of five different categories: driving force, pressure, state, impact and response. The goal in this PhD study was to improve the DPSIR framework in order to increase our understanding of the problems related to water quality. The adaptation boiled down to a conceptual change for 'state' and 'impact', which were adapted to mean 'physicochemical state', 'hydromorphological state' and 'ecological state' and the related impacts. This can be justified by the fact that according to the WFD the surface water state is at first hand defined by the ecological quality indicators (e.g. macroinvertebrates, fish and macrophytes), supported by physicochemical and hydromorphological quality elements. By proposing these changes in the DPSIR framework, this research considers the recommendations stated by Vanrolleghem (2010a) regarding the necessity of improving integrated assessment in model-supported river basin management.

The IEMF is 'integrated' in the sense that the output of the water quantity and quality models is the input for the ecological models. In the IEMF, dynamic (e.g. MIKE11 (DHI, 1999)) or steady state (e.g. QUAL2Kw (Pelletier et al., 2006)) models can be used for water quantity and quality simulations. Daily average data of physicochemical and hydraulic variables predicted with these models at each sampling station are used as input data for the ecological models, for model integration purposes, during the scenario analysis. This means that simulations based on data of hourly fluctuations of water quality and quantity variables in dynamic models are simplified by average conditions in this scenario analysis. This approach is considered valid in this PhD study, because aquatic macroinvertebrates have relatively long life cycles and are confined for most part of their life to one locality on the river bed. Macroinvertebrates integrate environmental conditions over longer periods of time (weeks, months, years) (Goethals, 2005). De Pauw and Hawkes (1993) pointed out that the biotic component of an aquatic ecosystem can be considered as the 'memory' of the ecosystem, integrating a wide range of ecological effects over time. In the IEMF, direct relations between a set of predictor variables (physicochemical and hydraulic) and ecological response variables (e.g. biological index value) are calculated, without incorporating feedback loops.

Once the integration of models is performed, they can be used for predicting the ecological water quality considering different simulation scenarios of river management. Thus, the IEMF allows considering the impact of driving forces such as the overflow of the sewer systems, the overload or shutdown of WWTP, the upgrading of WWTP and dam discharges on the ecological water quality. The IEMF allows assessing ecological degradation in rivers and streams, helps to understand this problem and could provide crucial information for water managers in environmental decision making. The integration of models through the IEMF allowed a holistic assessment that could not be achieved when looking at each individual component of the system separately (i.e. the impact of a WWTP effluent, on the receiving river and a dam).

The applicability of the IEMF as decision support tool in river water management and the integration of models towards the assessment of the ecological state of rivers will be shown in three case studies (Chapters 3-5) and discussed in Chapter 6. The proposed IEMF was applied on three rivers with different geographical locations, altitude, size and pollutions problems: (1) a deep lowland river in a tropical region, the Cauca river in Colombia (Chapter 3); (2) a shallow mountain river in a tropical region, the river Cuenca in the Andes of Ecuador (Chapter 4); (3) a lowland river in a temperate zone, the Drava river in Croatia (Chapter 5). Considering the limited information in the case studies in Colombia and Ecuador, only three of the four basic modelling components of the IEMF were implemented, a water quantity model, a water quality model and ecological models. In the case study of Croatia, IEMF links an integrated urban drainage system, considering the discharge of the WWTP, with the ecological state of the receiving river by following the conceptual elements of the framework.

In the IEMF there is no hydromorphological model implemented. However, a hydromorphological assessment based on a categorical variable called 'Type' that holds information on the hydromorphological structure of the water body was considered. Two categories or levels were defined for this variable: (1) hydromorphological favourable (value of one): natural bank structure, mixed bottom substrate, thin sludge layer, meandering, heterogeneous bank and bottom structure; and (2) hydromorphological unfavourable (value of two): artificial bank structure, tick sludge layer, straight waterway, homogeneous bank and bottom structure. This hydromorphological 'Type' variable was considered as input variable for the ecological models developed. Considering the limited hydromorphological information in the case studies in Colombia and Ecuador, the hydromorphological assessment based on a categorical variable called 'Type' was only performed in the case study of Croatia.

1.2 Problem definition

The impact of both climate change and human activities on biodiversity and ecosystems poses a serious and growing threat to sustainable development and protection of the environment. Human activities can have a multitude of different effects on rivers and streams, and it is difficult to identify those that have the biggest impact on the river ecology. Thus, there is a need for the development of practical tools, such as ecological models, providing accurate ecological assessment of rivers and species conditions. This should allow preserving habitats and species, stop degradation and restore water quality.

The ecological river water quality is mainly affected by two types of pressures: (1) hydromorphological disturbances and (2) physicochemical pollution. Model integration in water management allows analyzing these two types of pressures. Hence, some conceptual frameworks have been developed as an alternative towards an integrated ecological assessment: (1) the Driving forces–Pressures–Chemical and Ecological states-Response (DPCER) framework (Rekolainen et al., 2003); (2) the Species at Risk (SPEAR) framework (von der Ohe et al., 2009); (3) the Physical Habitat Simulation Model (PHABSIM) of the In-stream Flow Incremental Methodology (IFIM) framework (Bovee et al., 1998; USGS, 2001) and; (4) Driver–Pressure–State–Impact (DPSI) framework (Jähnig et al., 2012). These conceptual frameworks consider either the link between physicochemical variables and the river ecology or the link between the hydromorphological variables and the ecological water quality. However, these approaches do not consider the simultaneous effect of physicochemical pollution and hydromorphological disturbances on the ecological state of the receiving river. Therefore, an integrated modelling framework, such as the IEMF, that considers the concept of ecological state, defined in terms of the quality of the biological community and the hydromorphological and physicochemical characteristics is necessary.

1.3 General objective and scope

The proposed research aims to develop and to evaluate an integrated ecological modelling framework for decision support in river management. The specific research goal is to propose a decision support tool for analyzing driving forces, such as the discharge of treated or untreated wastewaters and habitat degradation caused by changes in the hydromorphological conditions, which change the ecological water quality. The scope of this research includes physicochemical pressures such as the discharge of wastewater treatment plants and hydromorphological pressures such as changes in water course, current velocity, water depth, riverbed sediment composition and bank structure. Moreover by integrating four types of models to simulate WWTP processes, water quantity, water quality and ecological aspects, this PhD study aimed to assess the effectiveness of different wastewater treatment/disposal strategies in three different case studies.

1.4 General methods

From a technical point of view, there are two approaches that can be implemented during the integration of models. The monolithic approach, which uses an over-all model including more or less detailed representations of subsystems and the modular approach, which uses existing models and combines them into an integrated model (Kraft, 2011). The former has the benefit of control in the model design and linkage, but requires longer development time (i.e. elaborated manageability), a deep knowledge of the system and profound software skills. Another disadvantage of this approach deals with transparency, because of the high risk of conceptual errors hidden deep in their code (Kraft, 2011). The latter approach saves on development time (i.e. there are a lot of models already available), can be easily extended and is flexible to be modified, but requires additional work to link up existing models (Lam et al., 2004).

In the context of integrated ecological modelling for riverine systems, the modular approach is the most popular (van Griensven et al., 2006, Pauwels et al., 2010; Jähnig et al., 2012; Boets et al., in press b) because: (1) it allows including (detailed) water quantity and quality models already available; (2) it can operate both, at the coarse and small river basin scale levels; (3) the ecological models are based on specific characteristics of the studied river. It can be argued that in this context, the modular approach can be convenient, if the aim is transparency, analysis and hierarchical description of various processes and system components instead of reusability and just connecting individual models. Examples of the implementation of the modular approach by coupling water quality and quantity models with river ecological assessment models are presented by Pauwels et al. (2010) and Boets et al. (in press b). Other authors reported the link of the two first models with species distribution models to predict the habitat suitability for selected species in riverine systems (e.g. van Griensven et al., 2006).

Considering the aim of this research and the advantages of using the modular approach in an ecological modelling context described by Voinov et al. (2004), the modular approach was adopted for the proposed IEMF. In this approach, ecological models based on data-driven modelling techniques were integrated with river water quality and quantity models (Chapters 3, 4 and 5) and a model that simulates the outflow of a wastewater treatment plant (only in the case study of Croatia, Chapter 5). A current trend is the integration of data-driven models with physically based models in an optimal way (Solomatine et al., 2008). The idea is to combine models of different types and which follow different modelling paradigms, thus constituting hybrid models. These types of hybrid models can combine some of the advantages and eliminate some of the disadvantages of the existing models (Jorgensen and Fath, 2011). This was the approach followed in this PhD study with the IEMF. One of the challenges for ecologist and hydroinformaticians in this respect is to ensure that data-driven models are properly incorporated into the existing modelling and decision support frameworks.

The integration of hydromorphological, physicochemical and ecological components in sub-models was performed in the three case studies in Colombia, Ecuador and Croatia. Ecological models based on data-driven modelling techniques allowed predicting the ecological water quality in terms of the presence of selected species of macroinvertebrates and the river state according to ecological water quality indices. The integration of water quality and quantity models such as MIKE11 (DHI, 1999), QUAL2Kw (Pelletier et al., 2006) and the River Water Quality Model No.1 (RWQM1, Reichert et al., 2001a) with habitat suitability and ecological assessment models was implemented in the IEMF.

Additionally, the RWQM1 was linked to a model to simulate processes in a wastewater treatment plant (WWTP) (Activated Sludge Model No. 2d (ASM2d); Henze et al., 2000), implemented in the simulation platform WEST (World wide Engine for Simulation, Training and Automation; Vanhooren et al., 2003).

MIKE11 (case study in Colombia) and QUAL2Kw (case study in Ecuador) models have both two different simulation modules integrated, a module for water quality modelling and another for water quantity modelling. MIKE11 evaluates dynamic flow conditions, therefore, it was considered as a hydrodynamic model whereas QUAL2Kw only evaluates steady state flow conditions and thus it was considered as a hydraulic model. For using the RWQM1 (case study in Croatia) it was necessary to implement and to integrate both, the hydraulic and the water quality model by using Matlab (Matrix Laboratory 7.10; MathWorks, 2010) applications. These two models were developed considering the Continuous Stirred Tank Reactor in Series (CSTRS) approach (Whitehead et al., 1979). A flow chart for model selection considering the type of water quantity and water quality models in the IEMF is presented in Fig. 1.2. The choice of the type of water quantity and quality models has to do with the variability of flow and physicochemical parameters considered for modelling purposes, the sampling frequency (e.g. hourly, daily or once per year) and the availability of data. A summary of the different modelling techniques implemented in the three case studies is presented in Table 1.1.

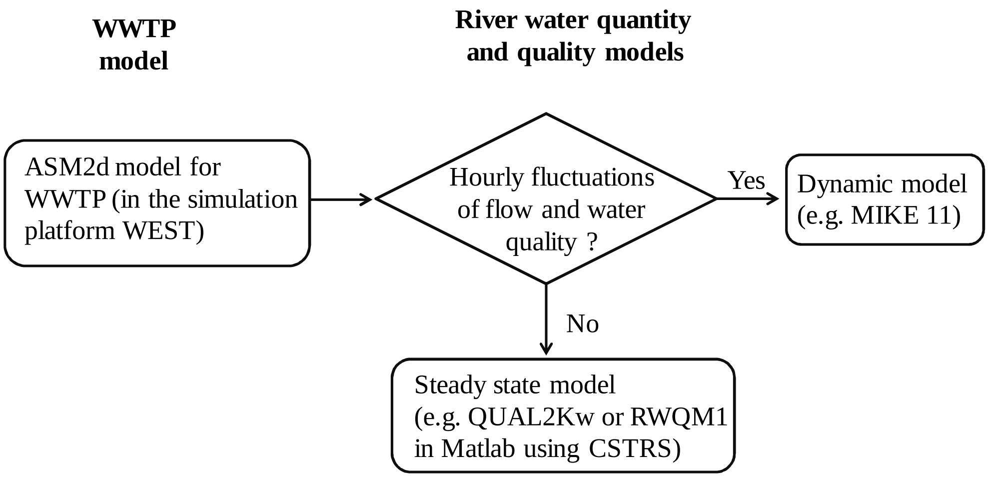


Fig 1.2. Flow chart for model selection in water quantity and quality modelling in the *Integrated Ecological Modelling Framework* (IEMF)

Table 1.1. Summary of the four type of models implemented in the *Integrated Ecological Modelling Framework* (IEMF) in the three case studies: (1) river water quantity model; (2) river water quality model; (3) wastewater treatment plant (WWTP) model; (4) ecological models (i.e. two types, habitat suitability model for selected species of macroinvertebrates and river ecological assessment model). (ASM2d: Activated Sludge Model No. 2d; RWQM1: River Water Quality Model No.1; CSTRS: Continuous Stirred Tank Reactor in Series; WEST: World wide Engine for Simulation, Training and Automation).

Country	River water quantity model	River water quality model	WWTP model	Ecological models
Colombia	MIKE 11 (dynamic model)	MIKE 11 (dynamic model)	---------	- Habitat suitability model - Ecological assessment model
Ecuador	QUAL2Kw (steady state model)	QUAL2Kw (steady state model)	---------	- Habitat suitability model - Ecological assessment model
Croatia	RWQM1 in Matlab using CSTRS (steady state model)	RWQM1 in Matlab using CSTRS (steady state model)	ASM2d for WWTP (in the simulation platform WEST)	- Ecological assessment model

(----): model not considered in this case study

For the ecological modelling, habitat suitability models for selected macroinvertebrate groups and ecological assessment models based on three different biological indexes based on macroinvertebrates were applied. These biological indexes were: (1) the Biological Monitoring Working Party index for Colombia (BMWP-Colombia; Zúñiga and Cardona, 2009); (2) the Biotic Integrity Index using aquatic invertebrates (IBIAP; Carrasco, 2008) in Ecuador and; (3) the Multimetric Macroinvertebrate Index of Flanders (MMIF; Gabriels et al., 2010) in Croatia. Multivariate statistics using Generalized Linear Models-GLM (i.e. logistic and negative binomial regression) and machine learning techniques (i.e. decision trees) were applied in these case studies. The choice of the ecological model type has much to do with the type of data (dichotomous (presence/absence), count data or continuous data) and availability of data. According to Vayssières et al. (2000), in case of small datasets (n = 30 records, considered in this PhD research), parametric methods such as GLM (e.g. LRM and NBRM), are generally more efficient than non-parametric methods such as decision tree methods (CT, RT and MT). A flow chart for model selection considering the type of ecological models in the IEMF is presented in Fig 1.3. More details about the materials and methods considered in this research are presented in Appendix A.

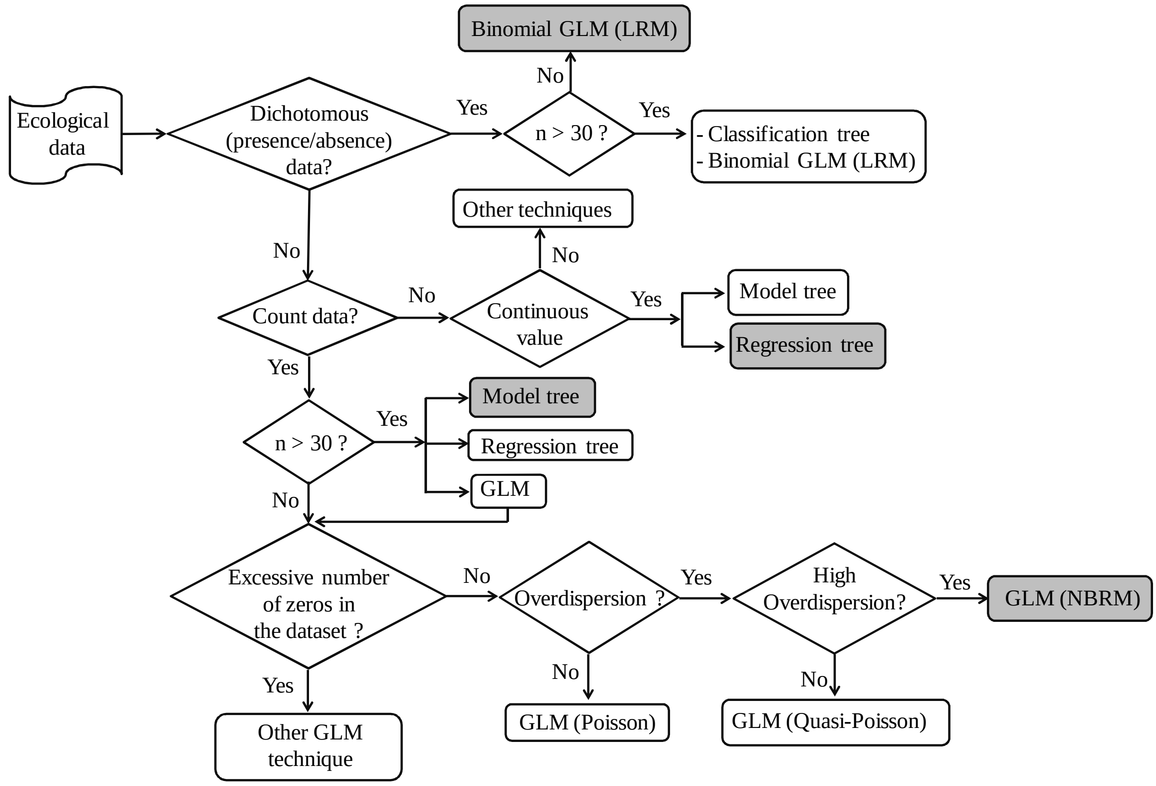


Fig 1.3. Flow chart for model selection in ecological modelling implemented in the *Integrated Ecological Modelling Framework* (IEMF). The modelling techniques used in this research are found in grey boxes. (LRM: Logistic Regression Models; NBRM: Negative Binomial Regression Models; GLM: Generalized Linear Model; n: number of samples in the dataset with simultaneous measurements of physicochemical, hydraulic/hydromorphological and biological information).

Two different types of model selection procedures were applied in the GLM techniques: (1) stepwise variable selection process with statistical considerations; (2) the multi-model inference based on the information-theoretic approach. Moreover, the classifier algorithms M5 (Quinlan, 1992; Wang and Witten, 1997) and M5P (Witten et al., 2011) were used for the regression trees (Breiman et al., 1984) and model trees (Quinlan, 1992) respectively. Additionally, to test the robustness of the models, different validation techniques were considered: (a) independent dataset validation, (b) internal validation, by resampling methods (Verbyla and Litvaitis, 1989), such as cross-validation and bootstrapping techniques. A summary of the different type of data-driven techniques implemented, the input and output variables, the software used, the model selection techniques and model validation procedures for the ecological models in the case studies is presented in Tables 1.2 – 1.4.

Table 1.2. Summary of the input and output variables considered for the ecological models in the three case studies. (DO: dissolved oxygen; BOD_5: five-day biological oxygen demand; ORGP: organic phosphorus; PO_4: phosphate; ORGN: organic nitrogen; NH_4^+: ammonia; NO_3: nitrate; D: water depth; V: water velocity; BMWP: Biological Monitoring Working Party; IBIAP: Biotic Integrity Index using aquatic invertebrates; MMIF: Multimetric Macroinvertebrate Index of Flanders).

Country	**Ecological models**	**Input variable (Predictor variables)**	**Output variable (Response variables)**
Colombia	- Habitat suitability model - Ecological assessment model	DO, D, V	- Ephemeroptera and Haplotaxida (presence/absence) - BMWP index (value between 0-120, count data)
Ecuador	- Habitat suitability model - Ecological assessment model	DO, temperature, BOD_5, FC, Flow, D, V	- Trichoptera and Physidae (presence/absence) - IBIAP index (value between 0-16, count data)
Croatia	- Ecological assessment model	DO, BOD_5, ORGN, $NH4^+$, NO_3, ORGP, D, V, hydromorphological type	- MMIF index (value between 0-1, continuous value)

Table 1.3. Summary of the main characteristics of the habitat suitability models and ecological assessment models in the three case studies. (LRM: Logistic Regression Models; NBRM: Negative Binomial Regression Models; MT: Model Trees; RT: Regression Trees; n: number of samples in the dataset with simultaneous measurements of physicochemical, hydraulic/hydromorphological and biological information; CCI: Correctly Classified Instances, *K*: Cohen's kappa coefficient, AUC: area under the receiver-operating-characteristic curve, *r*: Pearson correlation coefficient, R^2: determination coefficient, RMSE: root mean square error).

Country	**n**	**Ecological models**	**Type of data-driven technique**	**Type of variable**	**Model fitting (perfomance indicators)**
Colombia	15	- Habitat suitability model	LRM	Dichotomous (presence/absence)	CCI, *K* , AUC
		- Ecological assessment model	NBRM	Count data (0-120)	R^2, r
Ecuador	60	- Habitat suitability model	LRM	Dichotomous (presence/absence)	CCI, *K* , AUC
		- Ecological assessment model	MT	Count data (0-16)	R^2, r, CCI
Croatia	96	- Ecological assessment model	RT	Continuous value (0-1)	CCI, RMSE, r, R^2

Table 1.4. Summary of the software used, the model selection techniques and model validation procedures for the data-driven models implemented for the ecological modelling in the three case studies. (LRM: Logistic Regression Models; NBRM: Negative Binomial Regression Models, MT: Model Trees; RT: Regression Trees)

Country	Type of data-driven technique	Software used	Model selection technique	Model validation procedure
Colombia	LRM NBRM	R	Multimodel Inference based on AICc	Post-hoc evaluation of the model adequacy and predictive performance
Ecuador	LRM	XLSTAT	Stepwise based on likelihood ratio test with $p > 0.05$	3-fold cross validation
	MT	WEKA	M5P algorithm	
Croatia	RT	Matlab	M5 algorithm	- Independent dataset - Bootstrapping

Finally, simulations of scenarios were implemented to evaluate the impact of different river restoration plans, such as the upgrading of the existing wastewater treatment plants, on the ecological state of the receiving river.

1.5 Summary of content

The thesis research is divided in three core parts dealing respectively with:

- State-of-the-art of integrated ecological modelling of rivers and decision support in river management (with a focus on macroinvertebrates, Chapter 2);
- Application of the integrated ecological river modelling approach (with three case studies, Chapter 3, 4 and 5);
- General discussion and conclusions (Chapter 6). It includes some practical recommendations for integrated ecological modelling of rivers.

The chapters of the thesis are arranged as follows: Chapter 1 gives a general introduction; Chapter 2 presents a review of the state-of-the-art of integrated ecological modelling of rivers and decision support in river management, Chapters 3, 4 and 5 present three case studies of the application of the IEMF approach in rivers located in Colombia, Ecuador and Croatia respectively, and Chapter 6 presents a general discussion of the results, conclusions and recommendations for further research.

Chapter 2: State-of-the-art of integrated ecological modelling of rivers and decision support in river management

2.1 River water quality problems and water quality regulations

The misuse of freshwaters, rapid deterioration, scarcity and climate change pose a serious and growing threat to sustainable development and protection of the environment (Radif, 1999; Postel, 2000; Palmer et al., 2008). These problems will intensify unless effective and concerted actions are taken. Challenges remain widespread and reflect severe problems in the management of water resources in many parts of the world (Radif, 1999). The optimal balance between the different stakeholder activities needs a more in depth insight in the integrated water resources management (Molle, 2009).

One of the worldwide problems that affect the quality of water resources, has been controlled or uncontrolled discharges of wastes from agricultural, urban or industrial activities. These discharges can potentially affect human health and aquatic life, limit water use, affect riverine ecology and cause loss of amenity. River water quality assessment in many countries relies on physicochemical standards, however there is a gap concerning the impact of different pressures on river biota, which are used to assess river water quality. These pressures include physicochemical pollution (e.g. organic enrichment, eutrophication and acidification), physical changes and anthropogenic manipulation of the aquatic habitat (e.g. canalization, impoundment, river regulation). Nevertheless, international legislation such as the Water Framework Directive (WFD; European Commission, 2000), the Clean Water Act of 1972 and the Water Quality Act of 1987 (USEPA, 2011), changed the conventional practice by considering the importance of ecological assessments of receiving waters. During the last two decades, it has been emphasized that bio-monitoring of surface waters is a complementary tool for water quality assessment (European Commission, 2000; USEPA, 2011).

The WFD (European Commission, 2000), which aims to achieve a Good Ecological State of all European water bodies, introduced the integrated approach in river management, considering the concept of ecological state. Part of assessing the ecological state is monitoring the presence and diversity of aquatic species. However, species or diversity loss might have more than one possible cause. Therefore, in the WFD the ecological state is referred in terms of the quality of the structure and functioning of aquatic ecosystems, considering biological, hydromorphological and physicochemical quality elements. Moreover, the WFD promotes a combined water management of the legal emission limit values and the recipient quality standards and encourages the use of decision support tools such as water quality models. For these reasons, the development and use of water management tools for decision support, such as water quality, water quantity and ecological assessment models is necessary.

2.2 Integrated ecological river assessment

2.2.1 Hydromorphological quality

During the last two decades, the study of the effects of hydromorphological pressures on stream biota have been focusing on two main topics: (1) the identification of flow regimes for ecological protection (e.g. Stalnaker et al., 1995; Bovee et al., 1998; Hughes and Louw, 2010; Paredes-Arquiola et al., 2011; Jähnig et al., 2012) and; (2) the design and evaluation of river restoration schemes (e.g. Bockelmann et al., 2004; Tomsic et al., 2007; Everaert et al., 2013). Several river assessment studies based on hydromorphological characteristics have been developed, and they are mainly classified according to three approaches: broad scale assessment, microhabitat assessments and empirical habitat models (Maddock, 1999).

In the last decade, there was a gradually growing awareness that habitat variables, linked to the hydromorphological structure of the river play an import role in the ecological functioning of surface waters (Vaughan et al., 2009; Timm et al., 2011). The legislation put forward by the European WFD (European Commission, 2000) is an example of the aim of considering the hydromorphological elements in the ecological water quality assessment. The WFD uses the term hydromorphology to describe the hydrologic and geomorphic elements of river habitats. Important hydromorphological elements include: (1) morphology (including river sinuosity, water depth, water velocity, slope and river bottom substrate) and its variability; (2) the flow regime (including low flows, average flows and high flows, their timing, magnitude, frequency, duration) and; (3) weed cutting and dredging. The WFD aims at obtaining a ‘good ecological state’ of all water bodies in the European member states by 2015 (European Commission, 2000). Improving monitoring and assessment of habitat features of target species linked to river hydromorphology is a key aspect in water quality management of surface waters. The composition of macroinvertebrate communities is often linked to variables associated with stream hydraulics (Statzner and Higler, 1986; Kemp et al., 2000; Newson et al., 2012). Statzner et al. (1988) recommended that more complex hydraulic variables should be used, on top of the simple variables such as water depth and water velocity. Statzner and Higler (1986) suggested that measurements of water velocity, depth, substrate roughness, surface slope and hydraulic radius should be used in future hydraulic studies applied to benthic invertebrates (i.e. animals living at the bottom of a river). Furthermore, efforts have been done to establish an index to assess the hydromorphological water quality in function of the occurrence of different macroinvertebrate species (Kaeiro et al., 2011; Extence et al., 1999).

2.2.2 Physicochemical water quality

Water quality assessment can be defined as the evaluation of the physical, chemical and biological nature of water in relation to natural quality, human effects and intended uses. Historically, river management actions and research mainly focused on physicochemical water quality state as driver for ecological responses in river systems (Vaughan et al., 2009). Until 2000, this train of thought was considered as the core of river water quality assessment, research and management (European Commission, 2000; USEPA, 2011). The major source of organic and inorganic matter pollution includes the discharge of domestic and industrial wastewaters and agricultural discharges from livestock production, fertilizers and pesticides. The impacts generated by these pollution discharges includes the depletion of dissolved oxygen concentrations, the increase of organic and inorganic matter, eutrophication (nutrient enrichment) and contamination by hazardous compounds, that cause disturbances of the functioning of the ecological system. A decrease in the physicochemical water quality leads to loss in diversity of aquatic organisms and the disturbance in the ecosystem functioning (Chapman, 1996; Laws, 2000). Most processes in rivers are highly linked to each other and the change of one variable can lead to in-balance of many other quality variables. This domino-effect can lead to an irreversible deteriorated state of the river water quality.

Two categories of physicochemical river pollution can be distinguished. The first category is called point source pollution, which is a form of pollution concentrated at one point in space. The second category is called diffuse or non-point source pollution. Examples of point source pollution includes treated (controlled) or untreated (uncontrolled) discharges of industrial or urban wastewaters. Diffuse pollution includes different sources such as, runoff of fertilizers and pesticides from agricultural soils and rural residential developments. The assessment and control of non-point sources of pollution is more complex compared to point sources, because the effects of diffuse pollution both in time and space are difficult to quantify and to assess.

Historically, different organizations of several nationalities involved in water resources control have used physicochemical indices for water quality assessment. Nowadays, more environmental agencies, universities and institutes are turning to Water Quality Indices (WQIs) and Water Pollution Indices (WPIs) to facilitate interpreting physical, chemical and biological data. These indices lead to an evaluation of the water quality by means of a mathematical expression representing all evaluated variables. WQIs and WPIs reduce a great amount of physicochemical variables to a simple expression, to enable easier interpretation of monitoring data. The main difference between WQIs and WPIs include the form how they evaluate pollution processes and the number of variables taken into account in each formulation. A water quality index basically consists of a simple expression of more or less complex variables, which serve as water quality measurements. A number, a range, a verbal description, a symbol or a colour could be used to represent the index. Several WQIs and WPIs have been created based only on physicochemical variables in different countries, such as the WQI-NSF developed in 1970 by the National Sanitation Foundation of the United States and the Dutch Index in the Netherlands, among others. Regarding the WPIs, the Bacterial Pollution Index (BPI), the Nutrient Pollution Index (NPI), the Production Respiration Index (PRI), the Organic Pollution Index (OPI), the Industrial Pollution Index (IPI) and the Pesticide Pollution Index (PPI) were all developed in the Netherlands in the framework of the AMOEBA project (A General Method of Ecological and Biological Assessment) (Brink et al., 1991).

2.2.3 Biological water quality

Monitoring the quality of a freshwater ecosystem should not rely on physicochemical analyses alone. The discharge of wastewaters with organic, inorganic and toxic substances in rivers, changes the normal water quality and habitat conditions, affecting the biota composition and changing the occurrence of dominant species groups. Besides, higher mortality at any life stage, deformities and changes in the behaviour or metabolism have also been reported (Chapman, 1996). Those alterations are easily detected during biological monitoring, but are hardly detected during physicochemical monitoring (De Pauw and Vanhooren, 1983). Gabriels (2007) pointed out that biological monitoring and biological criteria provide the most robust approach to track the state of waters, because waterways that cannot support healthy biological communities are unlikely to support ecosystem services provided by these systems. Thus, biological monitoring can provide more information on the state of an ecosystem than physicochemical monitoring or hydromorphological assessment alone.

The biotic component of an aquatic ecosystem can be considered as an ‘integrating-information-yielding unit’ for assessment of its quality. Biological communities also integrate the effects of mixed types of stress and in certain cases already respond before analytical detection allows for. The advantages of biological monitoring are the limited equipment requested, the low cost and the large areas that can be evaluated in a short period of time. However, expert staff is necessary for the species identification (Chapman, 1996). Moreover, in order to obtain a complete evaluation of the aquatic community, different groups should be evaluated which can make it impractical and expensive (Metcalfe, 1989).

Hynes (1960) presented one of the best examples related to the impact of human activities, such as the discharge of wastewaters, on the river ecology. The diagrammatic representation illustrates the response of the river ecology in function of the physicochemical composition of the river water quality (Fig. 2.1). The concentration of different physiochemical components and the distribution of diverse organisms like bacteria, fungi, algae and macroinvertebrates are represented in the length profile of the river.

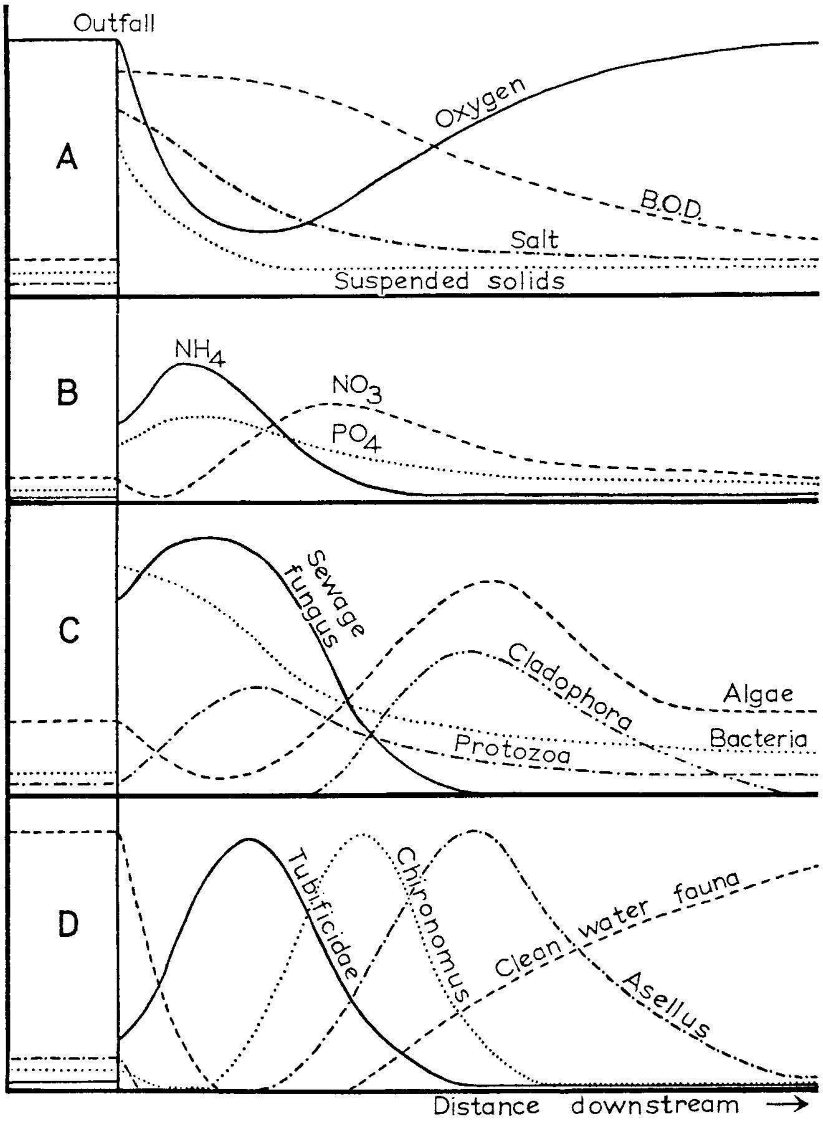


Fig. 2.1. Example of the effects of an organic effluent on the ecological state of the downstream river system. A and B represent the changes in physicochemical variables, C the change in number of micro-organisms and D the changes in the number of macroinvertebrates (Hynes, 1960).

Complementary, Vannote et al. (1980) introduced the River Continuum Concept (RCC) (Fig. 2.2), which provides an insight in the way biological communities may change from the headwater stream (source of the stream and smallest permanently flowing stream) to larger rivers in the absence of human influence. The RCC divides a river system into three major groups comprising headwater streams, mid-sized streams and large rivers (Gordon et al., 2004). According to the RCC, the biotic and abiotic structure and function of the running water is characterized by longitudinal, vertical and lateral gradients (Deksissa, 2004). For example, the RCC predicts that the number of species will increase and the proportion of shredders (organisms that consume leaf materials) will decrease from headwater streams to larger rivers. In mid-sized rivers there is a shift to grazer communities, and in the lowland reaches the collectors dominate. Moreover, as the size of a river increases from a headwater stream to a mid-sized river, the influence of the surrounding riparian forest decreases due to the change in the dominant biological community. The physical basis of the RCC is the size of the river or stream (stream order) and location along the stream. The stream order is an approximate measure of stream size and correlates with a number of other, more precise size measures including the area drained, volume of water discharged, and channel dimensions (Allan and Castillo, 2007). A large stream order corresponds to a larger stream. The smallest permanently flowing stream is referred to as first order. The union of two first-order streams results in a second-order stream, the union of two streams of second order results in a third-order stream, and so on.

The RCC summarizes expected longitudinal changes in energy inputs and consumers as one proceeds from a first-order stream to a large river. The RCC predicts that primary production will be lowest in forested headwaters (i.e. first-order streams), increase in more open, midsized rivers, and decline in turbid, higher-order stream segments (Vannote et al., 1980). A production to respiration ratio (P/R) approaching 1 indicates that much more energy to the food web is supplied by primary production within the stream channel. Thus, in first-order streams and higher-order stream segments, a low P/R indicates that the majority of the energy supplied to the food web derives from organic matter and microbial activity, and mostly originates as terrestrial production outside the stream channel (Allan and Castillo, 2007). An important upstream–downstream linkage is the export of fine particulate organic matter (FPOM) from the headwaters to locations downstream (Allan and Castillo, 2007).

The RCC has been widely used in river water quality assessment and modelling (e.g. Shanahan et al., 2001; Carpenter, 2001). However, there are two main weaknesses of the RCC (Gordon et al., 2004): (1) it only applies to perennial streams (a stream that has continuous flow all year round, during years of normal rainfall), and it does not account for disturbances that interrupt the natural pattern, such as dams and water diversions and; (2) the lack of consideration of movement of water onto floodplains during flood events.

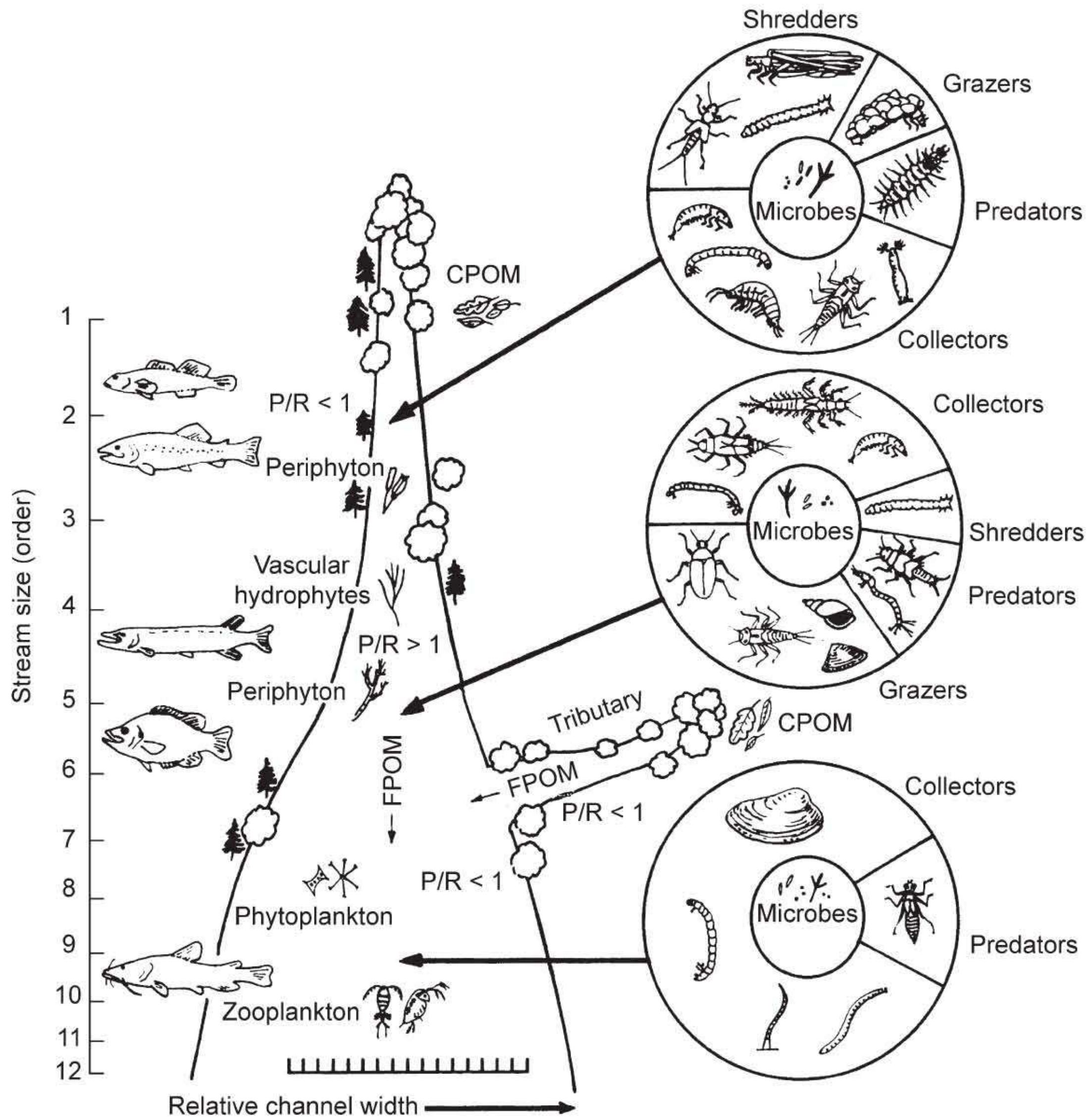


Fig. 2.2. River Continuum Concept (RCC): relationship between the stream size and the progressive shift of structure and function of river or stream communities. The relative proportions of various feeding groups are shown in the circles. (P/R: production to respiration ratio; FPOM: fine particulate organic matter; CPOM: coarse particulate organic matter) (Reproduced from Allan, 1995 after Vannote et al., 1980).

Among the biological communities, macroinvertebrates are by far the most frequently used group of bioindicators in standard water management, because they are ubiquitous and abundant throughout the whole river system and they play an essential role in the functioning of the river continuum food web (Goethals, 2005). They are visible to the human eye and relatively easy to sample and identify. Generally, macroinvertebrates are considered as those invertebrate animals inhabiting the aquatic environment that are large enough to be caught with a net or retained on a sieve with a mesh size of 250 to 1000 µm, and thus can be seen with the unaided eye. The majority of aquatic macroinvertebrates has a benthic life and inhabits the bottom substrates (sediments, debris, logs, macrophytes, filamentous algae, etc.). Other representatives of the macroinvertebrates, however, also serving as bioindicators, are pelagic and freely swimming in the water column, or pleustonic and associated with the water surface (Goethals, 2005).

Having relatively long life cycles and being confined for most part of their life to one locality on the river bed, aquatic macroinvertebrates act as continuous monitors, integrating water quality over a longer period of time (weeks, months, years) (De Pauw and Hawkes, 1993). They also constitute a taxonomically very heterogeneous group, showing a broad spectrum of responses to each form of stress, including physicochemical pollution (e.g. organic enrichment, eutrophication, acidification), and physical changes and anthropogenic manipulation of the aquatic habitat (e.g. canalisation, impoundment, river regulation). Macroinvertebrates can thus be used for the assessment of the water as well as the habitat quality and enable a holistic assessment of streams (Goethals, 2005).

However, the use of macroinvertebrates as indicators of river (water) quality has also limitations. Quantitative sampling for example is difficult because of their non-random distribution in the river bed. Because of the seasonality of the life cycles of some invertebrates, e.g. insects, they may not be found at some times of the year (Goethals, 2005). Therefore, having seasonal monitoring campaigns enables this seasonality to be taken into account when interpreting the data. Besides water quality, other factors such as current velocity, depth, nature of the substratum, water temperature and light penetration are also important determinants of benthic communities. Goethals (2005) pointed out that of these the related factors of current velocity and nature of the substratum are overriding ones determining the nature of the invertebrate community. Since these environmental conditions differ along the river in different zones, different communities become established at different sites with the same water quality. Therefore, in practice where possible, sampling sites having similar environmental conditions are selected or a typology is developed consisting of distinct river types with selected sampling and assessment systems (Goethals, 2005).

A last limitation of macroinvertebrates is their restricted geographic distribution, the incidence and the frequency of occurrence of some species being different in rivers throughout the region. Furthermore, because of their geographic distribution, species at the edge of their natural distribution range are theoretically more sensitive to additional stress – pollution than those at the centre of their distribution. It would therefore not be possible to have a universal system of biological assessment based on the response of the same species/taxa (Goethals, 2005).

2.3. Mathematical modelling

In general, there are two types of mathematical modelling approaches, called stochastic and deterministic modelling. In a stochastic model the outputs are not unambiguously determined by the model inputs. These types of models contain elements of randomness and the predicted values depend on probability distributions. Including randomness in a model can be considered in order to account for the uncertainty associated with the model input variables, parameter values and model structure (Deksissa, 2004). On the other hand, a deterministic model contains no elements of randomness or does not comprise uncertainty, thus the model output is a single value. A complex deterministic model is opportune only when the provided degree of detail is really necessary. A graphical representation of the differences between stochastic and deterministic models is presented in Figure 2.3. A stochastic model contains stochastic input disturbances and random measurement errors. If they are both assumed to be zero, then the stochastic model will reduce to a deterministic model provided that the parameters are not estimated in terms of statistical distributions (Jorgensen and Fath, 2011). A deterministic model assumes that the future response of the system is completely determined by knowledge of the present state and future measured inputs.

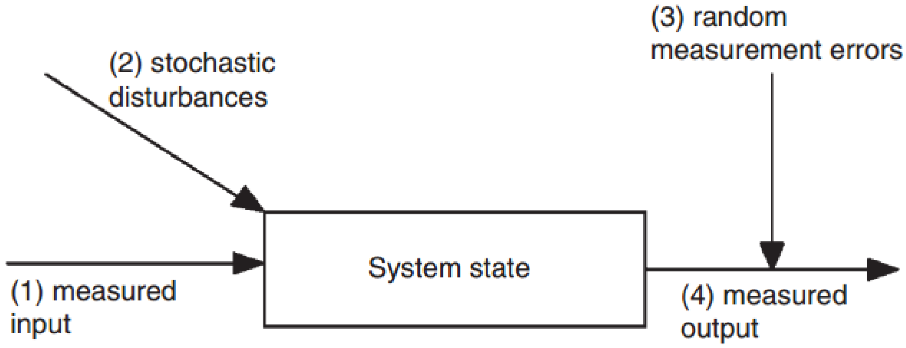


Fig. 2.3. Graphical representation of the differences between stochastic and deterministic models. A stochastic model considers (1), (2), and (3), while a deterministic model assumes that (2) and (3) are zero. Source: Jorgensen and Fath (2011).

Many of the parameters used in hydrological, hydraulic, water quality and ecological modelling are dependent on random forcing functions or on factors that cannot be included in our models without making them too complex. In these cases, it is recommended to apply stochastic models whenever the randomness of forcing functions or processes are significant (Jorgensen and Fath, 2011). By using Monte-Carlo simulations based on this knowledge, it is possible to consider the randomness (Jorgensen and Fath, 2011). By running the model many times, it becomes possible to obtain the uncertainty of the model results. Jorgensen and Fath (2011) presented some of the advantages of using stochastic models: (1) they are able to consider the randomness of forcing functions or processes and; (2) the uncertainty of the model results are easily obtained by running the model many times. The main disadvantages of this modelling approach are: (1) the distribution of the random model elements must be known and; (2) model implementation could have high complexity and require many hours of computer time.

A deterministic model can be further described as mechanistic (white-box), black-box and grey-box model. White-box (mechanistic) models are based on physical, biological and chemical laws, such as conservation of mass, momentum and energy, whereas the black-box (e.g. data-driven) models are those models that are not based on any physical or biological laws; instead they are based on data-driven transfer functions or processes (e.g. decision tree models, GLMs, Artificial Neural Networks-ANNs). If a model contains elements of both, white-box and black-box models, the model is called a grey-box model (expert knowledge-based models) (Adriaenssens, 2004). The approach that is preferred in a specific case depends on the aim of the research, the knowledge of the system processes and state variables in the system, the required properties of the model and the dataset available.

Data-driven models are useful in solving a practical problem or modelling a particular system or process if: (1) a considerable amount of high-quality data (reliable and relevant) describing this problem is available; (2) there are no considerable changes to the modelled system during the period covered by the model and; (3) there is little knowledge about the studied system. Such models are especially effective if it is difficult to build knowledge-driven simulation models (e.g. due to lack of understanding of the underlying processes), or the available models are not adequate enough (Solomatine et al., 2008). Data-driven models typically do not really represent the physics of a modelled process; they are just devices used to capture relationships between the relevant input and output variables. However, as Solomatine et al. (2008) stated, these models could be more accurate than process models since they are based on objective information (i.e. the data), and the latter may often suffer from incompleteness in representing the modelled process. On the other hand, mechanistic or expert knowledge-based modelling could be more appropriate when: (1) only a limited dataset or low-quality (unreliable or irrelevant) data are available; (2) there are considerable changes to the modelled system during the period covered by the model and; (3) there is a considerable knowledge about the studied system.

With regard to the temporal representation of the model, the distinction should be made between steady-state and dynamic (unsteady-state) models. In steady-state models, all inputs and state variables are constant in time. In dynamic models, however, input variables and state variables may vary with time, and thus result in a time variable output.

2.4 Major types of river system models

2.4.1 Hydraulic modelling

In general two methods can be used to simulate dynamic water movement (flood propagation) in rivers: the complex hydraulic routing method solving the 'St. Venant' equations (De St. Venant, 1871) and the conceptual hydraulic routing method (Deksissa et al., 2004). Generally, in the case of water quantity modelling of rivers, the 'St. Venant' equations, which include the continuity (mass balance) and momentum equations (momentum balance), are cross-sectionally integrated (1D). The form of a hydrodynamic model depends on assumptions made on characterizing turbulence. When the wind shear and eddy losses are omitted, the equations for a one-dimensional channel are as follows:

Continuity equation including lateral inflow (mass balance):

$$\frac{\partial Q}{\partial x} + \frac{\partial A_{cross}}{\partial t} = q \tag{2.1}$$

Momentum equation (momentum balance):

$$\underbrace{\frac{1}{A_{cross}}\frac{\partial Q}{\partial t}}_{\text{local acceleration}}+\underbrace{\frac{1}{A_{cross}}\frac{\partial}{\partial x}\left(\frac{Q^2}{A_{cross}}\right)}_{\text{convective acceleration}}+\underbrace{g\frac{\partial h}{\partial x}}_{\text{pressure force}}-\underbrace{g\left(S_o\right.}_{\text{gravity force}}\underbrace{\left.-S_f\right)}_{\text{friction force}}=0 \tag{2.2}$$

Kinematic wave

Diffusive wave

Dynamic wave

where, Q = flow rate [m^3s^{-1}]; A_{cross} = cross-sectional area [m^2]; h = absolute elevation of water level from the datum [m]; g = gravitational acceleration constant [m^2s^{-1}]; q = lateral inflow per unit length [m^2s^{-1}]; S_o = river channel side slope [-]; S_f = friction slope [-]; x = longitudinal distance of the river [m].

The ‘St. Venant’ equations require numerical methods (typically finite difference and finite element methods) to solve them. These methods require small time steps to overcome the numerical problem of instability. One-dimensional dynamic river water quantity models that are based on the full ‘St. Venant’ equations, which are included in software packages such as MIKE11 (DHI, 1999), CE-QUAL-ICM (Cerco and Cole, 1995) and DUFLO-EUTROF1 (Alderink et al., 1995). The full ‘St. Venant’ equations are rarely solved in water quantity and quality modelling practices because the solution of the equations tends to be complex and requires a lot of computational time. That is why Chow (1981) suggested simplification to these equations. Depending upon whether the flow is steady or unsteady (dynamic water movement) and which simplifications are made, many different forms and approximations to the ‘St. Venant’ equations are known. The momentum balance in equation 2.2 can be simplified by dropping terms (Chow et al., 1988). When the pressure and acceleration terms are dropped (i.e. only the friction term and gravity force are considered), the equation describes the kinematic wave only, which is limited to the monotonically decreasing of the riverbed. When the variation of flow is omitted (acceleration terms are dropped), the equation is simplified to the diffusive wave approximation, which allows describing backwater effects of weirs or other hydraulic controls like tidal effects. It can be applied when the river is not monotonically decreasing. If no term is ignored, the dynamic wave equations are able to describe the full dynamic wave.

In the absence of backwater and tidal effects caused by weirs or other hydraulic controls, such complex hydrodynamic model can be simplified into a conceptual hydraulic model in which the river is represented as a Continuous Stirred Tank Reactor in Series (CSTRS) (Whitehead et al., 1979; Beck and Reda, 1994; Deksissa et al., 2004). The CSTRS is a surrogate for the complex hydrodynamic model which combines the continuity equation with an analytical or empirical relationship between the storage of water in the system (or reservoir) and the outflow (Deksissa et al., 2004). An analytical way to express this relationship is by applying the Manning equation, whereas the empirical way establishes a relation between the outflow and storage by stage-discharge relationships. The concept of representing the river as a cascade of linear reservoirs has been applied by several authors (Camacho and Lees, 1999; Deksissa et al., 2004; Deksissa, 2004; Benedetti et al., 2007) in hydraulic modelling and it is linked to the concept of CSTRS. The results presented by these authors suggested that without sacrificing model simplicity, the CSTRS approach enables a good prediction of water movement compared with the full 'St. Venant' equations.

2.4.2 Physicochemical water quality modelling

To protect surface waters from all kind of sources of pollution, a holistic water quality regulation is required, such as the WFD (European Commission, 2000) or the American Clean Water Act of 1972 and the Water Quality Act of 1987 (USEPA, 2011). This legislation promotes a combination of legal emission limit values and the recipient Environmental Quality Objective/Standards (EQO/EQS). The EQO/EQS approach is based on the receiving water quality (immisssion) rather than the effluent water quality (emission) (Vanrolleghem et al., 1996). In an immission-based approach, mathematical models are required in order to predict the possible river water quality in response to emissions to the surface water, the hydrologic/hydraulic regime and the related transport processes and the physicochemical and biological processes (Bauwens, 2009). An integrated river water quality and quantity model therefore assists the water quality managers (authorities) to achieve a predefined water quality objective.

The challenge of using mathematical models in developing countries, such as Colombia or Ecuador, as a decision support tool to evaluate river water quality remediation options is well documented (Ongley and Booty, 1999). These countries have limited financial resources and an increasing deterioration of the water quality of their rivers, therefore, a prioritization of investments in sanitation infrastructure is necessary. Moreover, in these countries the impact of sanitation infrastructures (e.g. WWTP) is typically assessed considering the achievement of legal physicochemical quality standards, but ignoring the ecological water quality of the receiving river. Modelling requires substantial investment in reliable data, development of scientific capacity and a relatively sophisticated management culture that are often not found in developing countries (Deksissa, 2004). Nevertheless, the evaluation of the impact of basin management plans and pollution control and sanitation programs on the river water quality strategies require a mathematical model to predict the in-stream fate of pollutants as well as to estimate the likely effects that the resulting water quality may have on existing and potential water uses. Furthermore, the complex relationships between waste load inputs, and the resulting water quality responses in receiving water bodies are best described using mathematical models.

Two methods can be used to model river water quality: the complex pollutant transport routing, also known as advection-dispersion model (ADE model) and the conceptual pollutant transport routing (Deksissa et al., 2004). The ADE model method is based on the principle of conservation of mass of solutes and Fick's diffusion law. The ADE model represents the three governing processes in river systems (i.e. advection, diffusion, and reactions) by using a set of complex differential equations. Analogues to the ‘St. Venant’ equations, the ADE equations are rarely applied in their full form or in the three directions (longitudinal x, vertical y and lateral z) (Rauch et al., 1998). Hence they are often applied in a simplified form (1D). To solve the ADE equations numerically, they are coupled to the numerical solution of dynamic water movement in open channels, such as provided for the full ‘St. Venant’ equations (e.g. software MIKE11; DHI, 1999). However, in most cases, for water quality issues the acceleration terms in the momentum balance of the ‘St. Venant’ equations rarely play a significant role and the typical time scales are amplified by conversion processes. For these reasons, the diffusive (Rauch et al., 1998) and kinematic wave approaches are often a satisfactory approximation to simulate water movement in river water quality modelling. Thus, for water quality studies often the equation of steady, gradually variable flow is employed, which may be further simplified to the Manning equation as done in QUAL2E or QUAL2K models.

As the application of the full ‘St. Venant’ equations already requires long computation times, further extension of this model towards integrated water quality modelling will result in even more computation time. Consequently, an option is to use a conceptual mechanistic surrogate model for the sake of faster simulation and easy implementation of water quality models (e.g. Meirlaen et al., 2001). The conceptual pollutant transport routing is based on the assumption that a natural water body can be represented by a cascade of CSTRS (Chapra, 1997). In the cascade of CSTRS approach, a water body is represented as one or more fully mixed tanks (stretches, applying a ‘box model’ (Shanahan et al., 2001). The concept of a cascade of CSTRS has been successfully applied in river water quality modelling by Deksissa et al. (2004); Deksissa (2004) and Benedetti et al. (2007).

As can be seen, several water quality models can be used in the IEMF, therefore, there is a need for setting up a technical base for standardised, consistent river water quality models and guidelines for their use. In this context, Vanrolleghem (2010b) presented a modelling guidance document to water managers and other interested stakeholders on the model-supported implementation of the WFD. An example of a development to address this topic is the RWQM1 produced by an International Water Association Task Team. The RWQM1 has the advantage compared with the MIKE11 and QUAL2Kw, that it was developed to be compatible with existing IWA Activated Sludge Models (ASM1, ASM2, ASM2D and ASM3; Henze et al., 2000). Therefore, the coupling of river water quality and WWTP models is better suited using the RWQM1. Shanahan et al. (2001) defined a six-step process to guide decisions on model structure applicable to the range of river conditions that fit the River Continuum Concept (Vannote et al., 1980). These are summarised as: Step 1: Definition of the temporal representation (dynamic compared with steady state) that focuses on transport terms of the model and requires listing of all characteristics time constraints of relevant processes; Step 2: Selection of spatial dimensions, including if and how the sediment is included in representation of the river; Step 3: Determine representation of mixing, which depends on step 2 and number of dimensions to be modelled. Whether modelled as dispersion or diffusion, the representation of mixing varies depending on the hydrometrics of the site; Step 4: Determine representation of advection which, like step 3, does not depend on the characteristics of the conversion processes and can be, indeed, modelled independently of the water-quality; Step 5: Selection of the biochemical submodels, and their reaction times. This step is treated in more detail in

Reichert et al. (2001) and Vanrolleghem et al. (2001); Step 6: Determine boundary conditions, which is intrinsically linked to choice of model dimensions.

In general physicochemical water quality modelling includes two types of phenomena (Chapra et al., 2008): (1) model kinetics (e.g. dissolution, hydrolysis, oxidation, nitrification, denitrification, photosynthesis, respiration, excretion and death); (2) mass transfer (e.g. reaeration, settling, sediment oxygen demand, sediment exchange, and sediment inorganic carbon flux). The significance of different water quality processes varies depending on the case study considered. For instance in shallow rivers (e.g. mountain rivers) reaeration processes are highly important and are represented by high reaeration rates due to high turbulence and high flow velocities. Whereas in deep rivers (e.g. lowland rivers), characterized by low reaeration rates, associated to low flow velocities, settling processes are the overriding processes.

2.4.3 Ecological river assessment and river species distribution modelling

The application of models in ecology is almost compulsory if we want to understand the function of such a complex system as an ecosystem (Jorgensen and Bendoricchio, 2001). Ecological water quality modelling is an effective tool to investigate the ecological state of surface water resources (Goethals and De Pauw, 2001) including the self-cleaning capacity. This ecological river state depends on the actual and historical immission characteristics (i.e. the concentrations in the river; Willems and Berlamont, 2002), hydrologic/hydraulic regime and morphologic characteristics (Fig. 2.4). The immission concentrations in surface water are the result of emissions to the surface water, the hydrologic/hydraulic regime and the related transport processes and the physicochemical and biological processes that occur in the surface water (Bauwens, 2009).

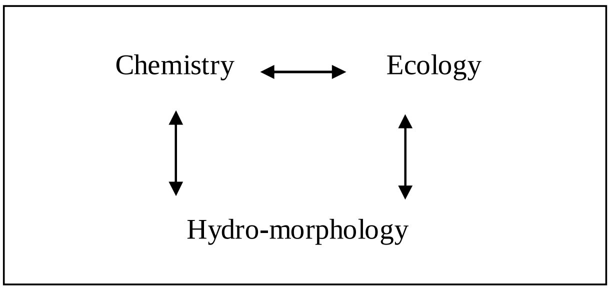


Fig. 2.4. Driving forces affecting the state of the surface water and representing the link between the different elements within the ecosystem.

The application of models in ecology is necessary if we want to understand the functioning of such a complex system as an ecosystem. However, the knowledge of ecological processes in ecosystems and the information available for a thorough insight into these processes have been much less developed and accessible compared with other science fields such as hydrodynamic or hydromorphological and physicochemical processes. In general, ecological modelling studies have three basic components: (1) a dataset describing ecological indices or the occurrence and/or abundance of the species of interest (response variables) and a dataset of explanatory variables (predictor variables); (2) a mathematical model that relates the species data to the explanatory variables and; (3) assessment criteria of the utility of the model developed in terms of a validation exercise or an assessment of model robustness (Rushton et al., 2004). In the following sections, two types of ecological models (mathematical models) that allow to understand the functioning of aquatic ecosystems are described. The first model is used to assess the ecological water quality of rivers (section 2.4.3.1) and the second model is used to predict river species distribution (section 2.4.3.2).

2.4.3.1 Ecological river assessment

Several researchers have used ecological models to support river management and water policy (Irvine et al., 2002; van Griensven et al., 2006; Deltares, 2009; Pauwels et al., 2010; Everaert et al., 2012; Everaert et al., 2013), mainly in the context of the European Water Framework Directive. However, there are still several knowledge gaps, emphasizing the need for the development of practical tools providing accurate ecological assessments of river and species conditions. This should allow preserving habitats and species, stop degradation and restore water quality. According to Goethals (2005), ecological models have several interesting applications in the context of river management and water policy. Firstly, through these models a better interpretation of the river state can be possible, the causes of the state of a river can be detected and assessment methods can be optimised. Secondly, these models can allow for calculating the effect of future river restoration actions on aquatic ecosystems and supporting the selection of the most sustainable options. Thirdly, these models can help to find the major gaps in our knowledge of river systems and help to set-up cost effective monitoring programmes.

Ecological mechanistic models (i.e. food-webs) have been mainly applied on lentic ecosystems (i.e. stagnant waters or systems with very low water velocity; e.g. lakes, ponds, reservoirs and wetlands) and the prediction of phytoplankton, zooplankton, macrophytes and fish communities. Few examples of the application of mechanistic models for predicting macroinvertebrates in lotic ecosystems (i.e. moving waters; e.g. rivers and streams) have been reported (Abdul-Aziz, 2010; Schuwirth et al., 2011).

During the last decade, the use of multivariate (statistical) approaches based on data-driven modelling techniques such as decision trees and GLM in an ecological context have been widely reported (Vayssières et al., 2000; Segurado and Araujo, 2004; Pearson et al., 2006; Guisan et al., 2007; Meynard and Quinn, 2007). Multivariate approaches are more appropriate than univariate approaches for the analysis of aquatic habitat as they inherently consider the interrelation and correlation structure of the environmental variables (Ahmadi-Nedushan et al., 2006). Data-driven modelling techniques can be used to build models for complementing or replacing physically based models (i.e. mechanistic models). Data-driven techniques, such as decision tree models and GLMs, are therefore more suitable for predicting macroinvertebrates and biological indices associated to them. These approaches have proven their applicability to various water-related problems: (1) habitat suitability (e.g. Goethals, 2005; Boets et al., in press a); (2) ecological assessment (e.g. Pauwels et al., 2010); (3) management of invasive species (e.g. Boets et al., 2010; Boets et al., in press b); (4) flow regimes identification for ecological protection (e.g. Jähnig et al., 2012) and; (5) the design and evaluation of river restoration schemes (e.g. Everaert et al., 2013), among others. The comparison of decision trees and GLMs in an ecological context have been reported (Vayssières et al., 2000; Segurado and Araujo, 2004; Pearson et al., 2006; Guisan et al., 2007; Meynard and Quinn, 2007). Some advantages and disadvantages of using decision tree methods (non-parametric technique) instead of GLMs (parametric technique) are discussed by Vayssières et al. (2000) and Debeljak and Džeroski (2011).

2.4.3.2 River species distribution modelling

Ecological water quality modelling is a time and cost effective method to investigate the relationship between the environmental conditions (e.g. physicochemical and hydraulic conditions) and the occurrence of organisms inhabiting the river. Models able to predict the habitat requirements of organisms help to ensure that planned actions for river restoration meet the required effects for the ecosystems. Thus, modelling of species distributions has become necessary in many aspects of biology, ecology and biogeography. Aquatic habitat suitability models could constitute a useful tool for decision-making within the framework of water management and applied biology. These type of models serve three main purposes: (1) to predict the (probability of) occurrence, abundance or distribution of species based on relevant abiotic and biotic variables; (2) to improve the understanding of species-habitat relationships and; (3) to quantify habitat requirements in terms of environmental variables.

The classic approach of quantifying habitat consists of estimating local habitat suitability curves which rely on available knowledge regarding optimum range of abiotic conditions for the targeted aquatic species. These suitability curves are analytical tools used to represent preferences of different aquatic species for various instream variables (e.g. water velocity, water depth, type of substrate, cover). In general, the preference curves are in the range of 0 to 1 for each variable with 0 meaning no preference for the particular habitat condition and 1 meaning maximum preference for the particular condition. Generally, physical habitat is dependent on more than one variable and several suitability curves must be combined to define a composite suitability index, such as the habitat suitability index (HSI), which is the most commonly used index of habitat. Several assumptions are implicitly used in discussed composite indices (Ahmadi-Nedushan et al., 2006): (1) all variables are equally important to the growth and survival of the aquatic organisms; (2) all environmental variables are independent and there is no interaction between them (Beecher et al., 2002). The first assumption can be relaxed by using the weighted product equation to consider the relative importance of each habitat variable to the aquatic organisms. However, HSI models have been criticized for the fact that they do not consider the interrelation and correlation structure of the habitat variables (Jowette, 2003; Leclerc et al., 2003). Moreover, in some cases, suitability curves are applied to climatic and geographical conditions different from those where they were developed. Habitat selection by macroinvertebrates is undoubtedly a multivariate process where location is selected based on several interacting variables (De Pauw et al., 2006). Therefore, the use of multivariate approaches such as decision trees and GLM allow taking into account the interaction between physicochemical and hydromorphological variables and to determine species response to cumulative effect of a number of environmental characteristics (Ahmadi-Nedushan et al., 2006).

2.5 The need for an integrated data collection and an integrated ecological modelling approach for decision support in river management

Assessment of the effect of human activities on river ecosystems requires indicators relating the cause to the effect (Fig. 2.5). Therefore, a cause–effect chain is distinguished whereby human disturbance changes abiotic steering variables, which in turn affect the biotic structural and functional characteristics of the river ecosystem (Lorenz et al., 1997). International legislation such as the WFD (European Commission, 2000), the Clean Water Act of 1972 and the Water Quality Act of 1987 (USEPA, 2011) emphasized the importance of integrated data collection in water quality assessment, considering hydromorphological, physicochemical and biological quality elements.

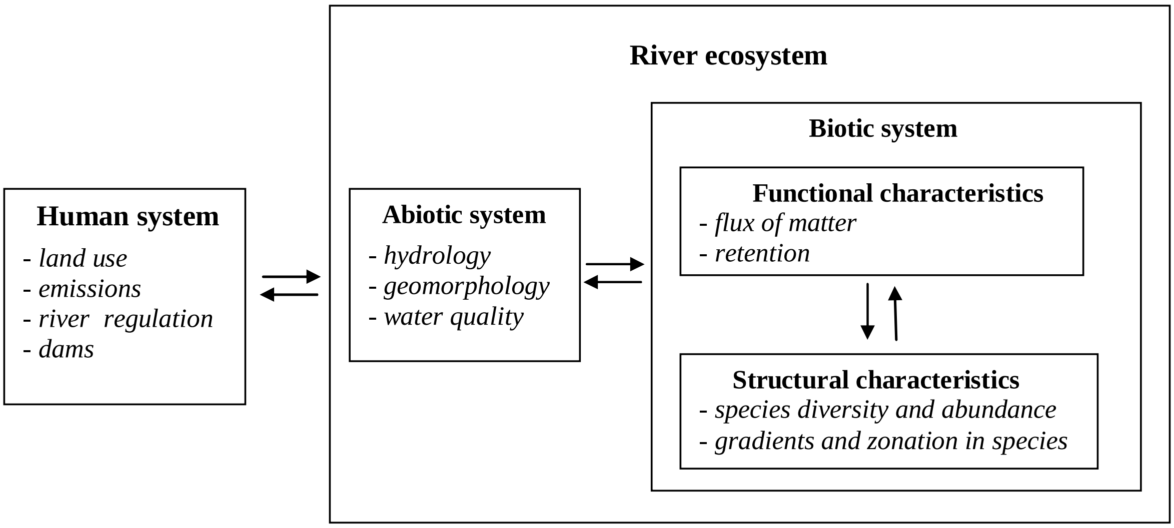


Fig. 2.5 Cause–effect chain of the human system influencing the river ecosystem Source: Lorenz et al. (1997).

When multiple impacts (e.g. habitat degradation and water pollution) are present, it is important to have a river monitoring strategy towards collection of integrated data. Therefore, linking environmental characteristics with community structure at each river reference site by using a defined set of variables and a combination of target groups representing the main functional levels of the ecosystems is required. Most often, a suite of macroinvertebrate criteria has been used in water quality assessment (Gore et al., 2001). Macroinvertebrates were investigated in this study, because they are good indicators for quality assessment in running waters, their sampling and identification is relatively simple and they are sedentary and have relatively long live cycles (De Pauw and Hawkes, 1993; Goethals, 2005). Furthermore, they play a key role in stream ecosystems, due to their intermediate position in the food chain linking allochthonous/autochthonous production with higher trophic levels (Munn and Brusven, 1991). Macroinvertebrates show a broad spectrum of responses to each form of stress, including physicochemical pollution and physical changes due anthropogenic manipulation of the aquatic habitat. Therefore, different factors besides physicochemical water quality are also important determinants of benthic communities (Fig. 2.6). Thus, biotic and diversity indices based on macroinvertebrates are used for identifying the water and habitat quality of streams and for measuring stresses to the environment.

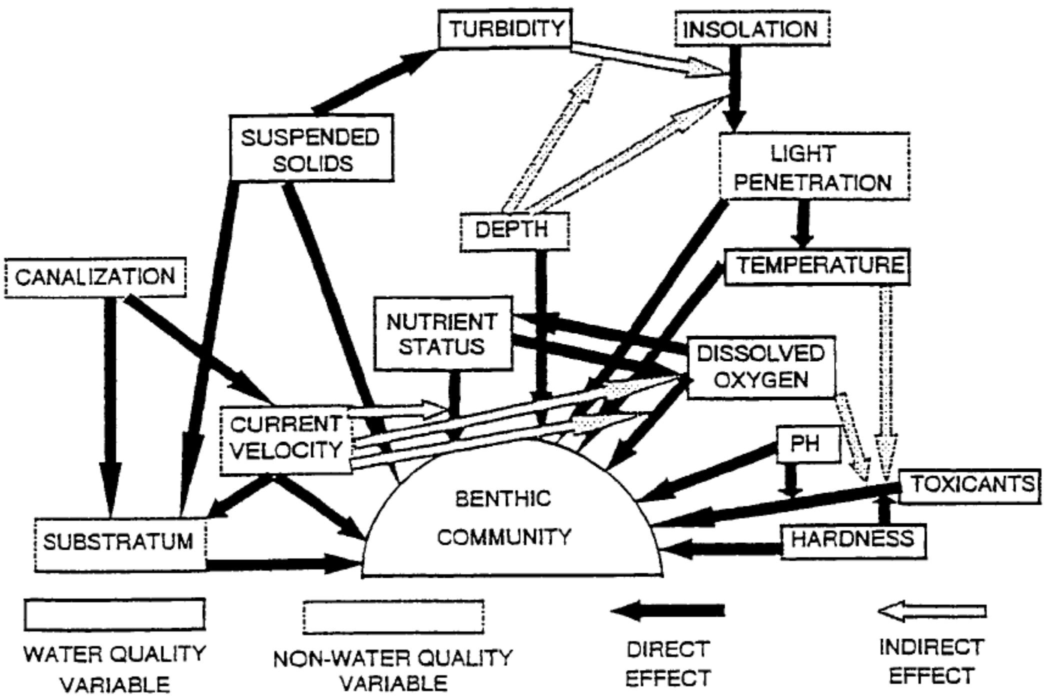


Fig. 2.6. Example of physicochemical and hydromorphological water quality determinants of benthic communities in rivers. Source: De Pauw and Hawkes (1993).

Robust ecosystem analysis of water resource systems remains elusive. An important reason is the difficulty to link engineering models used to simulate hydromorphological or physicochemical processes associated with project design or operation with ecological models used to simulate biological community attributes. The impact of the measures proposed in basin management plans and pollution control and sanitation programs on the river water quality is not straightforward, so it is unclear which combination of measures is most effective. Moreover, the impact of physicochemical pollution on the river ecology is significantly influenced by local conditions of current velocity, substratum and channel morphology (i.e. hydromorphological conditions). Therefore, it is necessary to link the environmental characteristics and the biological community structure at each reference site by using a defined set of variables and a combination of target groups (e.g. macroinvertebrates) representing the ecological water quality. Aquatic ecological models can guide management and policies and help in the design of monitoring programmes and interpretation of the results generated by such programmes. The use of appropriate mathematical models for surface water quality assessment can help to describe or to predict the impact of natural driving variables or anthropogenic pressures on habitat conditions and ecological processes and responses (at individual, population or community levels, Fig. 2.7).

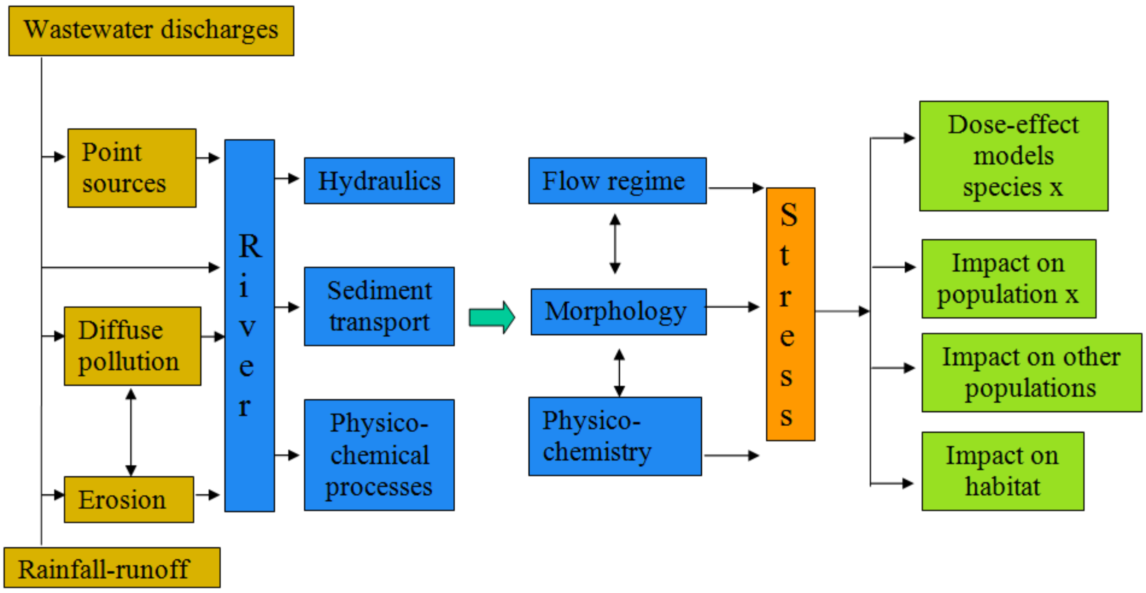


Fig. 2.7. General ecological modelling flow chart in a river. Source: adapted from Bauwens (2009).

Several attempts have been made to integrate hydromorphological, water quantity and quality models with habitat suitability and ecological assessment models based on macroinvertebrates, especially for Flanders (northern region of Belgium) and Netherlands (van Griensven et al., 2006; Deltares, 2009; Mouton et al., 2009a; Pauwels et al., 2010; Boets et al., in press b). However, the transferability of the ecological knowledge rules and data-driven models developed to other regions in the world is limited (Randin et al., 2006;

Fitzpatrick et al., 2007). Therefore, it is important to develop ecological models based on specific characteristics of the studied rivers. Thus, for the present study, the integration of models to evaluate the impact of wastewater discharges on the ecological water quality of rivers in Colombia (Chapter 3), Ecuador (Chapter 4) and Croatia (Chapter 5) is presented.

Considering the references and discussion presented in this section, it could be highlighted that physicochemical or hydromorphological evaluations should be always complemented by a biological assessment. Physicochemical or hydromorphological evaluations only reflect the condition of the river water quality at the moment the sample is taken (i.e. physicochemical monitoring) or when the hydromorphological pressures are assessed. However, these two evaluations do not indicate the effect on the biological community of the river (Cook, 1976; De Pauw and Vanhooren, 1983; Metcalfe, 1989). On the other hand, a river assessment based integrally on biotic indices is also incomplete. For instance, a poor biological score can be related to a combination of different physicochemical or hydromorphological conditions (Maddock, 1999).

2.6 Model uncertainty

Models are imperfect being a simplification of real systems and per definition, always contain errors in assumption, formulation and parameterization. Being simplified representations of the reality, the simulated (ecological) models can never be the same as the real nature, i.e. their results are somewhat uncertain. Uncertainty describes deviations between models' results and observed values. Uncertainty analysis implies the identification of errors, inexactness, imperfection and unreliability in the models. Uncertainty has different causes, including (Lek, 2007; Vaughan et al., 2009): (1) measurement errors (i.e. data imperfection); (2) the variability of models and parameters (models' sensitivity); (3) the lack of knowledge (i.e. limited scientific knowledge for some environmental processes); (4) conflicting evidence about a phenomenon and; (5) issues - especially in the future- that can never be known.

The importance of uncertainty in research and management has long been recognized, yet rarely addressed adequately (Vaughan et al., 2009). Uncertainty analysis should be included in modelling processes to avoid over-estimating confidence in conclusions or predictions, or setting unrealistic goals for management (Clark, 2002). Water quality management and river restoration projects provide good examples, being inherently complex and involving a high degree of uncertainty from a range of sources. Therefore, explicitly acknowledging uncertainties provides a way of managing unrealistic stakeholder, decision makers and societal expectations (Vaughan et al., 2009). Frameworks are required that consider uncertainty, along with tools with which to describe or quantify it (Clark, 2002). Many methodologies and tools suitable for supporting uncertainty assessment have been developed and reported in the scientific literature. Refsgaard et al. (2007) presented 14 methodologies to represent the commonly applied types of methods and tools for model uncertainty analysis. One of the most used methodologies to estimate uncertainty in hydrological, water quality and quantity models is the Monte-Carlo based method (Camacho and Lees, 1999; Camacho and González, 2008). Among other results, this method allows generating confidence bands for model results associated to 95% confidence intervals.

Regarding the ecological models, the possibility of selecting a confident set of models and making inferences derived from model averaging, when there is no single model that is clearly the best, shows the advantage of using GLMs compared with decision trees regarding model uncertainty assessment. There are three general approaches to assess model selection uncertainty using multi-model inference techniques (Burnham and Anderson, 2002): (1) theoretical studies, mostly using Monte-Carlo simulation methods; (2) the bootstrapping technique applied to a given set of data; and (3) using the set of Akaike's information criterion (AIC) differences (i.e., Δ_i) and Akaike model weights (w_i) from the set of models which fit to data. It is important to recognize that there is usually substantial uncertainty as to the best model for a given dataset. After all, these are stochastic biological processes, often with relatively high levels of uncertainty (Burnham and Anderson, 2002).

Yet, the shortcomings of using data-driven modelling techniques such as decision trees and GLMs for ecological modelling are acknowledged. These type of models should be used only in the range were they have been constructed. In order to preserve the statistical reliability and stability and to reduce uncertainty for example when performing simulations for future scenarios, extrapolation of the models outside their training range should be omitted (Araujo and Guisan, 2006). Moreover, data-driven models implicitly incorporate biotic interactions and negative stochastic effects that can change from one region to another. This can make models fitted for the same species, but in different areas and/or at different resolutions, difficult to compare (Guisan et al., 2002). For instance (Boets et al., 2013) found that, for a macroinvertebrate invasive species, extrapolation of logistic regression models developed with a dataset in Croatia applied on Belgium and vice versa seemed to be more difficult compared to classification tree models. Therefore, the application of these models is limited to the specific geographical area where they were developed.

Chapter 3: Case study 1: Integrated ecological modelling to analyze the impact of wastewater discharges on the ecological water quality of the Cauca river in Colombia

Adapted from:

Holguin-Gonzalez, J.E., Everaert, G., Boets, P., Galvis, A., Goethals, P.L.M. (2013). 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

Holguin-Gonzalez, J.E., Goethals, P.L.M. (2010). Modelling the ecological impact of discharged urban waters upon receiving aquatic ecosystems. A tropical lowland river case study: city Cali and the Cauca river in Colombia. In: Swayne, D.A., Yang, W., Voinov, A.A., Rizzoli, A., Filatova, T. (Eds.), 5th Biennial meeting of the International Congress on Environmental Modelling and Software (iEMSs 2010): Modelling for environment's sake, International Environmental Modelling and Software Society (iEMSs) Ottawa, ON, Canada. http://www.iemss.org/iemss2010/Volume2.pdf, pp. 1447-1455.

Chapter 3: Case study 1: Integrated ecological modelling to analyze the impact of wastewater discharges on the ecological water quality of the Cauca river in Colombia

Abstract:

In this chapter, the impact of wastewater discharges on the ecological water quality (EWQ) of the Cauca river in Colombia was investigated. The IEMF presented in Chapter 1, was tested in a deep lowland river in a tropical region. Two types of ecological models were developed, habitat suitability models for selected macroinvertebrate groups and ecological assessment models based on a macroinvertebrate biotic index (BMWP-Colombia). Four pollution control scenarios were tested. It was found that the foreseen investments in sanitation infrastructure will lead to modest improvements of the EWQ, with an increase lower than six units of the ecological index BMWP-Colombia. Advanced investments, such as the collection and treatment of all wastewater produced by the cities of Cali, Yumbo and Palmira and upgrading of the treatment systems should be considered to achieve a good EWQ. It was established that parametric methods such as Generalized Linear Models used in ecological modelling (e.g. logistic and negative binomial regression) are suitable for analysing integrated ecological data, which are characterized by small datasets, such as the one used in this study (n=15).

3.1 Introduction

The traditional management of sanitation infrastructure of urban wastewater systems aims at fulfilling the legal physicochemical quality standards, usually without taking into account the ecological state of the receiving waters. European legislation (Water Framework Directive (WFD), 2000/60/CE) changed the conventional practice by introducing the integrated approach in river management, considering the concept of ecological state. This state is specified in terms of the quality of the structure and functioning of aquatic ecosystems, considering ecological, hydromorphological and physicochemical quality elements. Moreover, the WFD promotes a combined water management of the legal emission limit values and the recipient quality standards and encourages the use of decision support tools such as water quality models. In the United States the importance of ecological assessments of receiving waters is postulated in the Clean Water Act of 1972 (CWA) and the Water Quality Act of 1987 (USEPA, 2011). During the last two decades, it has been emphasized that bio-monitoring of surface waters is a complementary tool for water quality assessment (USEPA, 2011). In developing countries, such as Colombia, a prioritization of investments in sanitation infrastructure is necessary due to the limitation of available financial resources and the increasing deterioration of the water quality. Therefore, in these countries, the development and application of integrated ecological modelling tools to support river management and water policy are necessary.

During the last decade, the integration of hydromorphological, physicochemical and ecological models for decision support in river management started gaining interest (Mouton et al., 2009a; Vaughan et al., 2009; Hughes and Louw, 2010; Boets et al., 2013). From an ecological point of view, benthic macroinvertebrates have been chosen as ecological indicators because they are expected to respond to both physicochemical and hydromorphological pressures, and can act as a link between primary producers and higher organisms (De Pauw and Hawkes, 1993; De Pauw et al., 2006). Recently, researchers emphasized in the integration of hydraulic/hydrodynamic models with habitat suitability indices (HSI) using habitat reference curves for macroinvertebrates (e.g. Bockelmann et al., 2004; Tomsic et al., 2007). This HSI approach considers hydromorphological pressures (e.g. changes in water depth, water velocity, type of substrate), but omits the impact of physicochemical pressures (i.e. physicochemical pollution). Jowette (2003) and Leclerc et al. (2003) criticized the use of these HSI models because these do not consider the interrelation and correlation structure of the habitat variables. Additionally, the transferability and applicability of these habitat suitability curves are limited, especially when they are being applied to different climatic and geographical conditions (Randin et al., 2006; Fitzpatrick et al., 2007; Strauss and Biedermann, 2007). More recently, Mouton et al. (2009a) considered the impact of these two types of pressures (i.e. hydromorphological and physicochemical pressures) on the ecological river quality, with an application of the Water Framework Directive Explorer (WFD-Explorer) toolbox. The WFD-Explorer includes a one-dimensional hydraulic model linked to a mass balance module that allowed them to predict the ecological water quality (EWQ) based on ecological expert knowledge rules. However, this toolbox simplifies water quality processes as a retention factor. Moreover, it operates at the coarse river basin scale level; whereas the impact of physical habitat changes on river biology occurs at smaller scale levels, such as mesoscale or microscale level (Mouton et al., 2009a). Additionally, the ecological knowledge rules implemented in the WFD-Explorer were developed based on empirical data of Dutch and Flemish lowland streams, therefore, the transferability of these rules to other ecoregions in the world is limited (Randin et al., 2006; Fitzpatrick et al., 2007).

Considering the limitations of the HSI and WFD-Explorer approaches, there is a need for an integrated approach that allows assessing simultaneously the impact of hydromorphological pressures and physicochemical pollution on the ecological river quality. This approach should include a detailed physical habitat and water quality model linked to ecological models based on specific characteristics of the studied river. Therefore, in this research the IEMF presented in Chapter 1, was tested on a case study of a lowland river basin in Colombia (Cauca river). In this study, three of the four basic modelling components of the IEMF (see Fig. 1.1, in Chapter 1) were considered. The first and second components, which correspond to river water quantity and quality modelling, were included in the MIKE 11 model (DHI, 1999). The third component included two types of ecological models: (1) habitat suitability models and; (2) ecological assessment models. This integrative framework was used to assess the ecological benefit of investments in sanitation infrastructure in the Cauca river by considering four pollution control scenarios.

The Environmental Authority in the Cauca Region (CVC) has been using a mathematical modelling approach since 1972 to support water management and to improve the water quality of the Cauca river. During the last decade (1997-2007), in the framework of the Cauca River Modelling Project (CRMP), the MIKE 11 model (DHI, 1999) was used to simulate the hydrodynamics and water quality of the river (CVC and Univalle, 2007). This modelling approach allowed getting insight into the processes that occur in the river under dynamic conditions, such as temporary variations of flows and polluting loads. However, the EWQ of the receiving river should be incorporated in this assessment, in order to guarantee the preservation of habitats and species, stop degradation and restore water quality.

3.2 Materials and methods

3.2.1 Study area

The Cauca river is the second most important river in Colombia and the main hydrologic resource of southwest Colombia. The Cauca river´s valley is especially important for the country's development and economy (CVC and Univalle, 2007). A significant part of the south-western manufacturing industry, the paper and sugar cane industry as well as part of the coffee producing zone are located along the river. The rapid urbanization and major economic development in the Cauca river´s valley, has led to dramatic degradation of the environment. There is an increasing deterioration of the water quality of this river due to wastewater discharges from domestic and industrial activities. This study focuses on the river stretch from the station Paso de La Balsa (abscissa 27.4 km and elevation of 965 meters above sea level-m.a.s.l) to the station Anacaro (abscissa 416.5 km and 805 m.a.s.l) (Fig. 3.1) with a total length of 389.1 km. Multiple water quality problems can be found in this zone, especially in the dry season, downstream from the cities of Cali, Yumbo and Palmira (main industrial cities in the region). Under low flow conditions the Biological Oxygen Demand (BOD_5) and Faecal Coliforms can rise up to 7.5 mg/L and $2.4*10^8$ MPN/100mL, respectively, whereas the Dissolved Oxygen (DO) concentration can drop near to zero mg/L. The city of Cali, with more than two million inhabitants, is the main source of pollution as 60% of all wastewater does not receive any type of treatment and is directly discharged into the Cauca river (CVC and Univalle, 2007).

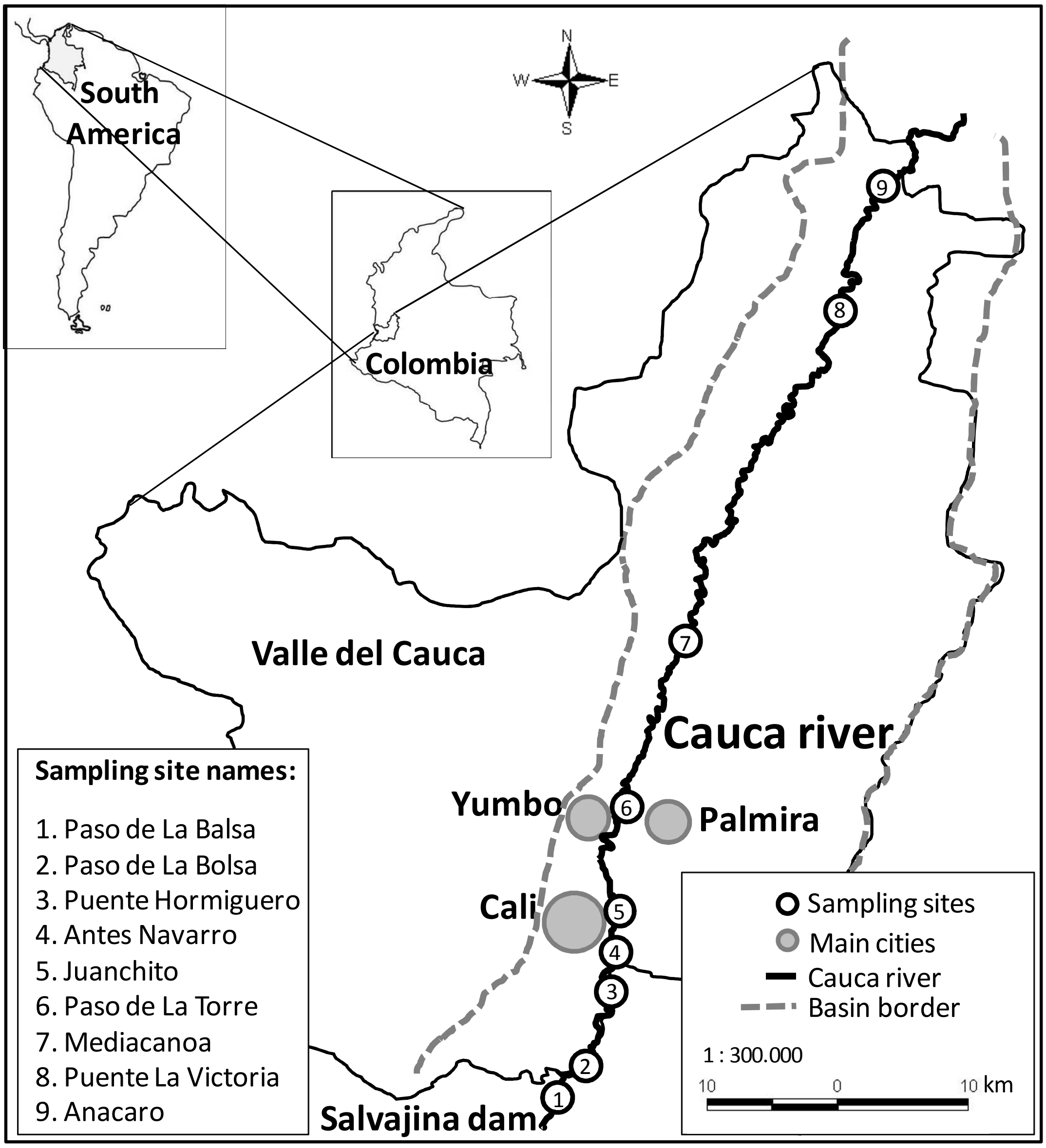


Fig. 3.1. Overview of the study area with indication of the Cauca river and the sampling sites in the Valle del Cauca region, Colombia.

3.2.2 Data collection, coupling of data and dataset pre-processing

The dataset used in this research corresponds to the information collected in a 10 year period (1996-2005) by the CVC and the CRMP Project in the Cauca river (CVC and Univalle, 2007). Two types of datasets were used, the first one for the implementation of the MIKE 11 model and the second one for building the ecological models.

Two monitoring campaigns with calibration and verification purposes of the MIKE11 model were carried out during August 2003 (low flow conditions) and February 2005 (high flow conditions) considering hourly measurements. These campaigns had a duration of respectively four (4) and five (5) days, a monitoring period between 12 and 24 hours per day, with a measuring frequency between 30 and 60 minutes for field parameters (flow, DO, temperature, conductivity and pH) and between six (6) and eight (8) hours for laboratory parameters (BOD_5, Chemical Oxygen Demand (COD) and Total Suspended Solids (TSS)).

For the ecological models, a dataset was developed which included simultaneous measurements (based on sampling location and time) of physicochemical data, hydraulic data and biological information. The biological information encompassed 32 records of macroinvertebrate communities from nine sampling locations collected between 1996 and 2004. At each sampling location (Fig. 3.1) the EWQ was assessed at least once in this period using the ecological index BMWP-Colombia (Zúñiga and Cardona, 2009). This index is calculated based on macroinvertebrate community composition and sensitivity to organic pollution and it is expressed as a value between 0 and 120; higher BMWP-Colombia scores reflect better river water qualities. The EWQ classes determined by this index were defined by Zúñiga and Cardona (2009): Class 1: very good EWQ (100 - 120); Class 2: good EWQ (61 - 99); Class 3: moderate EWQ (36 - 60); Class 4: deficient EWQ (16 - 35); Class 5: bad EWQ (< 15). Macroinvertebrate communities were sampled following the sampling protocol described by Zúñiga and Cardona (2009). Identification was carried out up to the required taxonomic levels, meaning family or genus level for all taxa (Zúñiga and Cardona, 2009). Unfortunately, some variables were not measured for one or more samples (incomplete measurement campaign). The dataset was, therefore, refined to ensure that the samples used in the analysis included measurements for all variables. This meant that 15 of the 32 sampling records were retained for analysis after coupling the physicochemical and hydraulic information with the biological data (see Appendix B; Table B.1).

MIKE 11 is a water quality model that predicts physicochemical variables under different water management scenarios. In order to enable the coupling between the ecological models and the MIKE 11 outcomes, only the six variables modelled by the MIKE 11 model (i.e. temperature, BOD_5, DO, flow, water depth and water velocity) and the biological information were retained. The final dataset for the ecological models consisted of these six variables (called predictor variables) and three response variables (presence/absence of two macroinvertebrate taxa and BMWP-Colombia values).

Two target macroinvertebrate taxa were selected for constructing the habitat suitability models, Haplotaxida (pollution tolerant taxon) and Ephemeroptera (pollution sensitive taxon). These two taxa are complementary ecological indicators, because their geographic distribution in the Cauca river (presence or absence) depends on their pollution tolerance (Zúñiga and Cardona, 2009), ranging from a tolerance score of 1 (very tolerant taxa) to 10 (most sensitive taxa). The pollution tolerance scores (PTS) for the family Tubificidae, which belongs to the Haplotaxida is one, whereas, the PTS for the Ephemeroptera families identified in this river (Leptohyphidae and Leptophlebiidae) lies between seven and eight (Zúñiga and Cardona, 2009).

The data available for building the ecological models were pre-processed considering three aspects: possible outliers, collinearity and relationships between the response variable and the predictor variables. Graphical tools, box plots and Cleveland dot plots were implemented to detect potential outliers (Zuur et al., 2007). Collinearity between the predictor variables was assessed by a Principal Component Analysis (PCA) and the Spearman rank (*S*) correlation coefficient. The *S* correlation coefficient was chosen rather than the Pearson correlation coefficient because the *S* correlation coefficient makes no assumptions about linearity in the relationship between the variables (Zuur et al., 2009). The correlation coefficients allowed exploring the correlation between the potential predictor variables. Based on the PCA and the correlation analysis different sets of predictor variables were tested for constructing the ecological models (see Appendix C1).

3.2.3 Water quality assessment

The water quality assessment of the Cauca river was performed considering the ecological and physicochemical water quality. In addition to the BMWP-Colombia (Zúñiga and Cardona, 2009), two physicochemical indices were considered, the Dissolved Oxygen Prati (DO-Prati) index (Prati et al., 1971) and an Expert Knowledge Based Index (EKBI) developed by the authors. Details about the water quality assessment of the Cauca river are presented in the Appendix C2.

3.2.4 Water quality modelling techniques

The three modelling components of the IEMF considered for this study were: (1) a river water quantity model, (2) a river water quality model and, (3) river habitat suitability and ecological assessment models. For the first and second components, the hydrodynamic and physicochemical water quality model MIKE 11 (DHI, 1999) was used. For the third component, logistic regression (presence/absence predictions) and negative binomial regression (BMWP-Colombia index predictions) were implemented. The selection of these two types of regression models is discussed further in section 3.2.4.2. Once the integration of models is performed, they can be used for model simulations. The ecological models developed were applied on the resulting hydraulic and physicochemical data of the water quality scenarios generated by simulations with the MIKE 11 model. An overview of the modelling techniques and different modelling processes implemented is presented in Table 3.1.

Table 3.1. Overview of the implemented modelling techniques, the different components of the model and the model building, validation, fitting and uncertainty (MSE: Mean Squared Error, CCI: Correctly Classified Instances, *K*: Cohen's kappa coefficient, AUC: area under the receiver-operating-characteristic curve, *r*: Pearson correlation coefficient, R^2: determination coefficient, LRM: Logistic Regression Model, NBRM: Negative Binomial Regression Model, GLUE: Generalised Likelihood Uncertainty Estimation).

Model component	Model building	Model validation	Model fitting	Model uncertainty
Water quality and quantity model (MIKE 11)	Constraint-based random search	Independent dataset	MSE	GLUE
Habitat suitability model (LRM)	Multi-model inference	Post-hoc evaluation of the model adequacy and predictive performance	CCI, *K*, AUC	Confident set of models
Ecological assessment model (NBRM)			r, R^2	

3.2.4.1 River water quantity and quality model

The hydrodynamic and physicochemical water quality model MIKE 11 (DHI, 1999) used in this research is a mathematical simulation model which was calibrated and verified for dynamic flow conditions in the framework of the CRMP Project (CVC and Univalle, 2007). The implementation of a simulation model begins with the representation of the characteristics of the system that are required to model. In the case of river modelling this representation corresponds to hydromorphological characteristics and the definition of the frontiers of the model (external and internal frontiers). The external frontiers correspond to the monitoring stations located upstream and downstream of the study stretch. The internal frontiers correspond to tributary rivers, water extractions and (wastewater) discharges. The MIKE 11 model was implemented by considering 62 cross sections, 2 external boundaries (monitoring stations Paso de La Balsa and Anacaro), 85 internal boundaries which include 38 rivers and streams, 9 municipal wastewater discharges, 12 industrial wastewater discharges and 37 water extraction sites. Each internal boundary was represented like a lateral discharge or extraction. The water quality modelling of the Cauca river was performed in the Level one of the MIKE 11 model, which includes temperature, BOD_5 and DO as state variables.

The monitoring campaign of 2005 (high flow conditions) was selected for calibration of the MIKE 11 because it included more wastewater discharges monitored and it had a longer monitoring time (5 days). Once a simulation model is calibrated, it should be validated using data obtained for water quality and hydraulic conditions different from those used for the calibration. By using the same calibration parameters, the model should have the capacity to reproduce the values of the new dataset. Thus, the validation of the MIKE 11 was performed with the monitoring campaign of 2003 (low flow conditions). The results of the calibration and validation of the MIKE 11 model can be analysed in two ways. The first analysis considers hourly variation of the physicochemical variables in each station during the monitoring days and the second analysis consists of an instantaneous profile of the values of the variables in all the stations simultaneously. This study focused on the first analysis, which gives a better idea of the modelling output under dynamic conditions.

A sensitivity analysis, based on the parameter perturbation method (Chapra, 1997), was performed to select the most sensitive calibration parameters. The re-aeration formula proposed by O´Connor and Dobbins (1956) gave the best correlations with the experimental re-aeration rates obtained during the CRMP Project. For the calibration a constraint-based random search method (Oddi et al., 2005) was implemented. For this method, thousands of combinations of the most sensitive calibration parameters (kinetic rates), considering values from uniform distributions, were evaluated with simulations considering the data of the monitoring campaign of 2005. The uniform distributions were estimated for each calibration rate parameter (A_i) considering the minimum (A_{min}) and maximum (A_{max}) values, reported by Bowie (1985) and Chapra (1997) and considering experimental values estimated in the CRMP Project (CVC and Univalle, 2007). By using a random function that generates a number between zero and one in the following equation, each value of the uniform distribution had the same chance to be selected in one of the thousand simulations:

$$A_i = A_{\min} + (A_{\max} - A_{\min}) * RAMDOM \tag{3.1}$$

The goodness of fit considered during the calibration was the Mean Squared Error (MSE). The MSE was calculated for each model run performed during the calibration and for each modelled variable. The model with the lowest MSE for the two variables (BOD_5 and DO) simultaneously was selected, leading to the best combination of values of the most sensitive calibration parameters. For the validation process the model was run using the data from 2003 without changing the calibrated parameters. Additionally, uncertainty analysis was performed using the concepts of the Generalised Likelihood Uncertainty Estimation methodology (GLUE; Beven and Binley, 1992), based on the results of the constraint-based random search method.

3.2.4.2 River habitat suitability and ecological assessment models

The approach followed for the ecological modelling in this research was to use multivariate statistics based on Generalized Linear Models (GLM). Parametric methods such as GLM are generally more efficient on small datasets than non-parametric methods such as Generalized Additive Models (GAM) or classification trees (Vayssières et al., 2000). GLM provide users with a conventional mathematical function and are better suited for analyzing ecological relationships, which can be poorly represented by classical

Gaussian distributions (Zuur et al., 2007). Considering these aspects, it was decided to implement two GLM techniques, logistic regression models (LRM) for predicting occurrence of macroinvertebrates and negative binomial regression models (NBRM) for predicting the value of the BMWP-Colombia.

LRM are the most frequently used approach of the GLM techniques for predicting the probability of species occurrence or distribution (Aspinall, 2002, 2004; Rushton et al., 2004; Ahmadi-Nedushan et al., 2006). LRM estimate the probability of a response variable (presence/absence) given a set of explanatory (predictors) environmental variables (e.g. DO, BOD_5). The BMWP-Colombia score is a non-negative integer value (count data) which ranges between 0 and 120, therefore, the GLM should be fitted with another type of distribution (non-Gaussian) such as Poisson, quasi-Poisson or negative binomial regression. In this research all three type of models were evaluated, however, the NBRM were finally implemented because the data were ‘‘overdispersed’’ and the plotted residuals did not show any trend (Zuur et al., 2009). This NBRM allow performing an ecological assessment by predicting the BMWP-Colombia value based on abiotic water quality variables (physicochemical and hydraulic variables). In order to enable the coupling between the ecological models and the water quality/quantity model, the regression models were developed exclusively with the variables modelled by the MIKE 11 model. Details about the implementation of the LRM and NBRM are presented in Appendix A.

The next step in the model building process is to identify the key explanatory variables for the LRM and the NBRM. Thereby, a multi-model inference technique based on the information-theoretic (I-T) approach (Burnham and Anderson, 2002), was coded in the software R (R Development Core Team, 2009). Details about the multi-model inference technique implemented are presented in Appendix C3. In the I-T approach inferences can be made from more than one model, something that cannot be done using the traditional model selection approach or the null hypothesis approach (Johnson and Omland, 2004). The second-order Akaike’s information criterion corrected for small sample size (AICc, Hurvich and Tsai, 1989) was used in this research for model selection. The relative probability of each model being the best model was calculated considering their Akaike weights (w_i). When no single model is overwhelmingly supported by the data (i.e. w_i max = 0.9), then a (weighted) model-averaging approach can be used (Gibson et al., 2004). This situation occurs because a number of models in the set may only slightly differ in their data fit, as defined by an information criterion. The advantage of the I-T model averaging procedure is that it accounts for model selection uncertainty to obtain robust variable estimates or predictions (Grueberg et al., 2011). Technical details about the full-model averaging approach are described by Symonds and Moussalli (2011).

For defining sets of “best models” in the I-T approach, two criteria were considered: a threshold value of AICc differences between models (Δ_i) and model performance. It is recommended that the set of “best models” have Δ_i values lower than four (Burnham et al., 2011) and good model performances (Symonds and Moussalli, 2011). To assess the predictive performances in the LRM three criteria were evaluated: 1) percentage of Correctly Classified Instances (CCI); 2) Cohen's kappa coefficient (*K*; Cohen, 1960) and; 3) area under the receiver-operating-characteristic (ROC) curve called AUC. More details and the physical meaning of these criteria are presented in Appendix A. For the threshold-dependent criteria (CCI and *K*), a cut-off value for species presence was based on the percentage of the samples in which Ephemeroptera and Haplotaxida taxa were present in the dataset (40% and 60% of the samples, cut-off of 0.4 and 0.6 respectively; Willems et al., 2008). In order to reach a satisfactory model performance in an ecological context, it is recommended CCI values higher than 0.7, *K* values higher than 0.4 and AUC values higher than 0.7 (Manel et al., 2001; Gabriels et al., 2007). To assess model performances in the NBRM the Pearson correlation (r) and the determination coefficient (R^2) were evaluated.

For the validation of the GLM models, post-hoc evaluation of the model adequacy (Zuur et al., 2009; Fox and Weisberg, 2011) and predictive performance of the selected models were implemented (Gibson et al., 2004). A sensitivity analysis, based on the parameter perturbation method (Chapra, 1997), was performed to quantify the effects of the input variables on the ecological models. Two types of procedures were implemented. In the first one, each input variable was increased or decreased by 10% and all other variables were kept fixed at the average value of the dataset, and the condition number was estimated for each parameter. In the second method, each input variable varied between the minimum and maximum values reported in the dataset, and all other variables were kept fixed at the average value.

3.2.5 Simulation of pollution control scenarios

The LRM and NBRM were used to make predictions about the dependent variables (i.e. presence/absence of macroinvertebrates and BMWP-Colombia values) based on other independent data. A total of four scenarios generated by simulations with the MIKE 11 model were evaluated (Table 3.2). The physicochemical and hydromorphological simulations of each scenario were used as input variables for the LRM and NBRM. Daily average predictions of these input variables at each sampling station were considered (the validation of this approach was discussed in section 1.1 in Chapter 1.).

Table 3.2. Description of the four different pollution control scenarios considered in this research (BOD_5: five day biological oxygen demand).

Scenario		Year	Projection of the average pollution load in the study area (ton/day of BOD_5)			Effective removal percentage in the scenario = R/P (%)
No.	Name		P: Produced	R: Removed	D: Discharged	
1	Current situation	2005	387.6	183.6	204.0	47.4
2	No investment	2015	511.3	256.6	254.7	50.2
3	Intermediate situation	2015	511.3	339.5	171.8	66.4
4	High investment	2015	511.3	404.9	106.4	79.2

The scenarios were developed for the year 2005 as a reference situation and the year 2015 as projected time period. The year 2005 was considered as reference situation because the Environmental Authority in the Cauca Region (CVC) started a sanitation program in that year and they wanted to evaluate the impact of the program after 10 years (year 2015). The sanitation program plans pollution control measures in the Cauca river basin and includes investments in wastewater treatment plants and clean technologies (CVC and Univalle, 2007). The reference situation (i.e. scenario for year 2005) considered low flow conditions (i.e. flow < 180 m^3/s in the Juanchito station) and it had detailed information about pollution loads and water quality of the Cauca river in the year 2005. For Juanchito sampling station, the CVC and Univalle (2007) reported a range of flows for dry (< 180 m^3/s), average (180 to 292 m^3/s) and wet conditions (>292 m^3/s).

In the framework of the CRMP project a total of 27 scenarios were run and the three most representative scenarios (with 2015 as projected time period) were selected for this study. The four scenarios (i.e. reference situation and three projected scenarios) considered the same river flow characteristics, which means that all considered dry season conditions (low flow), when critical conditions for the dilution of the pollution are observed. Thus, the change in the physicochemical variables (e.g. DO and BOD_5) was only related with the pollution control measures proposed in each scenario and was not influenced by a change in the dilution capacity of the river. Projections of average pollution load produced (P), removed (R) and discharged (D) by cities (e.g. Cali, Palmira and Yumbo) and industrial activities (e.g. paper, sugar cane and food industries) were calculated (ton/day of BOD_5) for each scenario in the study area (Table 3.2).

A projected time period of 10 years was used to consider the impact of the increase of the wastewater pollution load, due to the growth of the population and the industrial activity in the study area (CVC and Univalle, 2007). Additionally, the investments planned for the same time period, for collection and treatment of wastewater and clean technologies, were considered for each type of pollution source (domestic or industrial wastewaters). Moreover, the effective removal percentage, calculated as the ratio between the removed and produced pollution load was reported for each scenario (CVC and Univalle, 2007). The slightly higher value of the effective removal percentage of the scenario of no investment (scenario 2) compared with the current situation (scenario 1), is related with the increase of the pollution load which is removed in the wastewater treatment plant (WWTP). The pollution load removed increased from 183.6 ton/day of BOD_5 (scenario 1) until 256.6 ton/day of BOD_5 (scenario 2). This rise is related with the increase of wastewater due to the population growth in districts where there was already a sewer system connected to the WWTP. The projections of average pollution load discharged to the Cauca river were finally used as input information for the simulation of the scenarios using the water quality model implemented (MIKE 11).

3.3 Results

3.3.1 Water quality assessment and river water quality modelling

The ecological assessment of the Cauca river showed that BMWP-Colombia values were concentrated only in three EWQ classes: class 3 = moderate EWQ (moderately polluted); class 4 = deficient EWQ (polluted) and class 5 = bad EWQ (heavily polluted). The sensitivity analysis allowed identifying the most important calibration variables in the MIKE 11 model to predict BOD_5 and DO: the re-aeration rate (k_2), the carbonaceous organic matter degradation rate (k_1) and the sediment oxygen demand (SOD). The constraint-based random search method performed for the calibration process and uncertainty assessment was focused on these three variables. An example of the results of the calibration process of the model for DO and BOD_5 at a specific sampling station (Juanchito) considering dynamic conditions can be seen in Fig. 3.2a and 3.2b. The GLUE technique allowed generating confidence bands for the model results, the higher the confidence band, the higher the uncertainty of the model results. The model performance indicator MSE obtained during the calibration of DO and BOD_5 indicates that for the monitoring stations Hormiguero, Juanchito and Mediacanoa the minimum MSE value was 0.4, whereas for the rest of the stations, Puerto Isaacs and Paso de La Torre, the minimum MSE values were 0.9 and 0.8 respectively.

3.3.2 River habitat suitability and ecological assessment models

Regarding the collinearity analysis, the first two principal components (PCs) explained 83% (Spearman correlation coefficient) of the variance in the data. The first PC included temperature, flow, water depth and water velocity, whereas the second PC included BOD_5 and DO. Variables such as BOD_5 and DO ($S=-0.71$), temperature and water velocity ($S=-0.76$), flow and water depth ($S=0.62$) and flow and water velocity ($S=0.42$) were highly correlated. DO ($S=-0.76$) and BOD_5 ($S=-0.54$) had the highest correlation with the BMWP-Colombia. In order to avoid highly correlated variables and model overfitting, only DO, water velocity and water depth were kept as predictor variables.

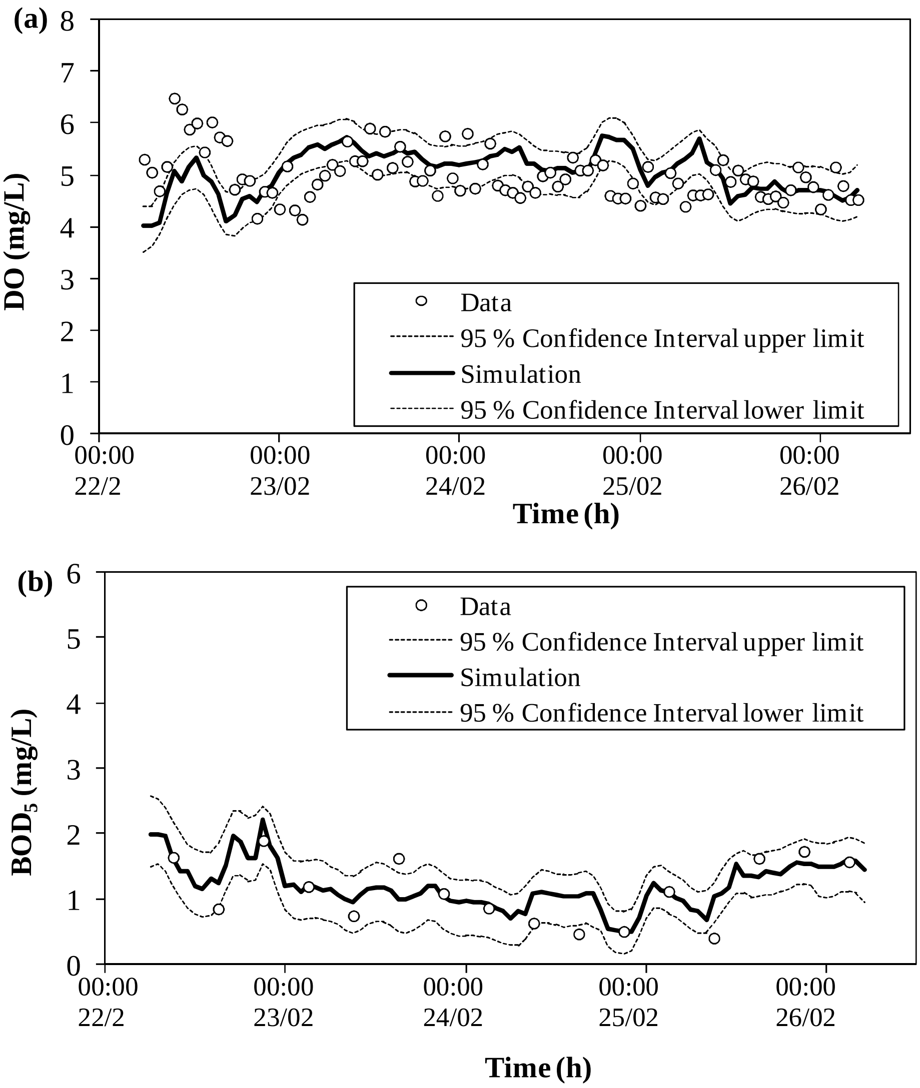


Fig. 3.2. Results of the calibration of the Cauca river water quality model at the station Juanchito for (a) dissolved oxygen (DO) and (b) five-day biological oxygen demand (BOD_5). Simulation period: 22–26 February 2005. Condition: High flows.

The AICc values, Akaike weight model rankings and performance criteria for all the LRM and NBRM are shown in Table 3.3. In this table the LRM and NBRM considered are ranked according to their AICc differences (Δ_i), from best (lowest AICc) to worst (highest AICc). The analysis of the set of "best models" showed that for Ephemeroptera predictions, the first five LRM had Δ_i lower than four units and good model performances (CCI>0.7, *K*>0.4 and AUC>0.7). This set of "best models" represents the 95% confidence set of models (CSM) (see cumulative Akaike weights in Table 3.3). For Haplotaxida predictions, the set of "best models" was conformed by the first three LRM, with good model performances and represented the 85% CSM. This indicates that these LRM correctly discriminate between occupied (presence) and unoccupied (absence) sites of these two macroinvertebrate taxa in the dataset. For the BMWP-Colombia predictions, the first six NBRM had Δ_i lower than four units, however, some of these models had very low performances compared with the others (third and fifth NBRM in Table 3.3). Therefore, it was decided to eliminate these two NBRM from the set of "best models", leading to a set of four "best models" with a range of moderate model performance (r = 0.61-0.69 and R^2 = 0.37-0.48), which represents the 80% CSM.

Given there is no single model that is clearly the best (i.e. w_i max = 0.9), a good approach is to acknowledge this model uncertainty and make inferences based on model averaging. Therefore, a model averaging by summing the Akaike weights was carried out on the set of models which represent an approximate 95% certainty (95% CSM). The average model for the LRM showed a very good performance with CCI=0.87, K=0.72 and AUC=0.94 for Ephemeroptera and CCI=0.80, K=0.59 and AUC=0.89 for Haplotaxida. The averaged model for the NBRM showed a moderate performance with r=0.69 and R^2=0.48. The values of the coefficients for the average model with the unconditional standard errors (i.e. non conditional of only one model) are presented in Table 3.4. Additionally, the relative importance of each predictor variable in the 95% confidence set of models is presented in this table. DO and water depth were the most important predictors for Ephemeroptera and the BMWP-Colombia, whereas DO was the most important for Haplotaxida.

The results of the post-hoc evaluation of the model adequacy based on diagnostic plots and the lack-of-fit test in the validation of the LRM and NBRM, are presented in the Appendix C4-C6. As an example of this analysis, the most important types of residuals defined in the GLM models, the Deviance and the Pearson residuals are presented for the most parsimonious model (lowest AICc). Neither outliers nor high-leverage points nor influential observations were identified. The dispersion parameter (Φ) for the Poisson regression model in the most parsimonious model was eight. The second alternative (NBRM) did not show any trend in the residual plots and was therefore selected to predict the BMWP-Colombia. The results of the sensitivity analysis of the ecological models are presented in Appendix C7. These results confirm those obtained with the I-T approach and showed that DO and water depth were the most important input variables (highest condition number) for the prediction of Ephemeroptera and the BMWP-Colombia, whereas DO was the most important input variable for the prediction of Haplotaxida.

Table 3.3. Results of the AICc-based model selection for the logistic regression model (LRM) and negative binomial regression model (NBRM) (Δ_i: AICc differences, w_i: Akaike weights, Cum. w_i: cumulative Akaike weights, CCI: Correctly Classified Instances, K: Cohen's kappa coefficient, AUC: area under the receiver-operating-characteristic curve, r: Pearson correlation coefficient, R^2: determination coefficient). The set of "best" LRM NBRM with AICc differences (Δ_i) lower than four units and good or moderate model performances are showed in bold. Good model performances in LRM are represented by CCI>0.7, K>0.4 and AUC>0.7, whereas moderate model performances of NBRM are represented by r = 0.61-0.69 and R^2 = 0.37-0.48.

Nr.	Model[a]	AICc	Δ_i	w_i	Cum. w_i	CCI	K	AUC
LRM for Ephemeroptera								
1	**D + DO**	**16.97**	**0.00**	**0.56**	**0.56**	**0.87**	**0.72**	**0.93**
2	**DO**	**20.02**	**3.05**	**0.12**	**0.68**	**0.73**	**0.44**	**0.85**
3	**D + V + DO**	**20.17**	**3.20**	**0.11**	**0.79**	**0.87**	**0.72**	**0.94**
4	**D**	**20.75**	**3.78**	**0.08**	**0.88**	**0.73**	**0.44**	**0.76**
5	**V + DO**	**20.86**	**3.89**	**0.08**	**0.95**	**0.80**	**0.57**	**0.85**
6	D + V	22.49	5.51	0.04	0.99	0.87	0.72	0.87
7	V	25.09	8.12	0.01	1.00	0.60	[b]	0.48
LRM for Haplotaxida								
1	**DO**	**18.72**	**0.00**	**0.59**	**0.59**	**0.87**	**0.72**	**0.87**
2	**V + DO**	**21.55**	**2.83**	**0.14**	**0.73**	**0.87**	**0.72**	**0.91**
3	**D + DO**	**21.90**	**3.17**	**0.12**	**0.85**	**0.87**	**0.72**	**0.87**
4	V	23.10	4.37	0.07	0.91	0.67	0.29	0.69
5	D + V	24.37	5.65	0.03	0.95	0.67	0.24	0.72
6	D + V + DO	24.91	6.19	0.03	0.97	0.87	0.72	0.87
7	D	24.93	6.21	0.03	1.00	0.60	0.00	0.54
NBRM for BMWP-Colombia		AICc	Δ_i	w_i	Cum. w_i	r	R^2	
1	**DO**	**127.83**	**0.00**	**0.38**	**0.38**	**0.61**	**0.38**	
2	**D + DO**	**128.44**	**0.61**	**0.28**	**0.66**	**0.67**	**0.45**	
3	D + V	130.83	3.01	0.08	0.74	0.56	0.31	
4	**V + DO**	**130.87**	**3.04**	**0.08**	**0.82**	**0.61**	**0.37**	
5	D	131.20	3.38	0.07	0.89	0.34	0.12	
6	**D + V + DO**	**131.31**	**3.48**	**0.07**	**0.96**	**0.69**	**0.48**	
7	V	132.26	4.43	0.04	1.00	0.19	0.04	

[a] Model includes variables: D, water depth; V, water velocity; DO, dissolved oxygen

[b] No possible calculation (division by zero)

Table 3.4. Model-averaged coefficients and relative importance of the predictor variables in the logistic regression model (LRM) and negative binomial regression model (NBRM) (S.E.: Standard error) in the 95% confidence set of models.

	Model averaged					
	LRM for Ephemeroptera		LRM for Haplotaxida		NBRM for BMWP-Colombia	
Variable	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
Intercept	1.177	4.021	4.399	2.735	2.946	0.580
Depth	-1.285	1.143	0.031	0.226	-0.094	0.105
Velocity	-1.035	3.399	-0.890	2.586	0.200	0.684
DO	1.030	0.865	-0.754	0.437	0.140	0.068

	Relative importance		
Variable	LRM for Ephemeroptera	LRM for Haplotaxida	NBRM for BMWP-Colombia
Depth	0.79	0.16	0.52
Velocity	0.20	0.26	0.24
DO	0.91	0.89	0.84

3.3.3 Integrated ecological modelling and scenario assessment

Profiles of average concentrations of DO and BOD_5 at the Cauca river were made for each pollution control scenario considering the results obtained with the MIKE 11 model (Fig. 3.3). Additionally, the impact of the different scenarios on the EWQ, expressed as the presence/absence of the two target species of macroinvertebrates and the value of the BMWP-Colombia index, was evaluated (Table 3.5 and Fig. 3.4a). Furthermore, the EKBI developed in this research and the DO-Prati index were applied (Fig. 3.4b and 3.4c).

The application of the integrated ecological modelling showed that the LRM and NBRM predicted the ecological impact well for the scenarios of pollution control in the Cauca river basin. In the scenario with high investment for pollution control (Table 3.5) an improvement of the EWQ is achieved, represented by the absence of Haplotaxida (pollution tolerant taxon) in the stations Nrs. 8 and 9 and the increase of the BMWP-Colombia (stations Nr. 5-9). On the other hand, in the scenario without investments for pollution control a deterioration of the EWQ is observed, represented by the absence of Ephemeroptera (pollution sensitive taxon) and the presence of Haplotaxida in the station Nr. 5, and the decrease of the BMWP-Colombia values (stations Nrs. 3-5 and 7-9). When the scenario of water quality objectives proposed by the government and the CVC is considered (intermediate situation), a limited EWQ improvement is achieved. There is absence of Haplotaxida in sampling station Nr. 8 and the increase of the BMWP-Colombia is limited to a smaller stretch (stations Nrs. 6-9) compared with the scenario with high investment (Table 3.5 and Fig. 3.4a).

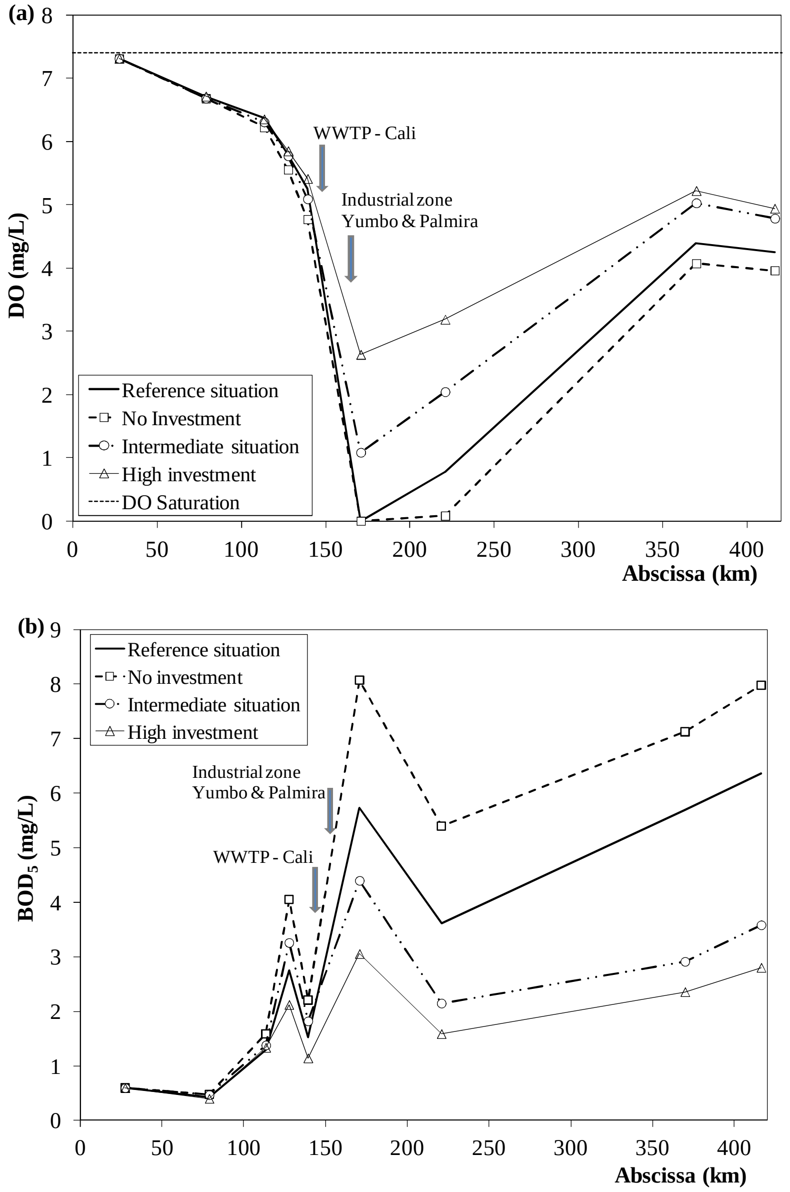


Fig. 3.3. Dissolved Oxygen (DO) and five-day biological oxygen demand (BOD_5) predictions in the Cauca river for the reference conditions and the four pollution control scenarios explained in Table 3.2. The arrows indicate the presence of a WWTP (wastewater treatment plant) or an industrial zone.

Table 3.5. Impact of different pollution control scenarios on the ecological water quality of the Cauca river, expressed as the presence/absence of two target species of macroinvertebrates and the BMWP-Colombia value.

Current situation scenario (Year 2005)

Sampling site		presence (1) or absence (0)		BMWP-Colombia
Nr.	Name	Ephemeroptera	Haplotaxida	Value
1	Paso de La Balsa	1	0	43
2	Paso de La Bolsa	1	0	42
3	Puente Hormiguero	1	0	43
4	Antes Navarro	1	0	30
5	Juanchito	1	0	28
6	Paso de La Torre	0	1	14
7	Mediacanoa	0	1	17
8	Puente La Victoria	1	1	32
9	Anacaro	1	1	28

No investment scenario (Year 2015)

Sampling site		presence (1) or absence (0)		BMWP-Colombia
Nr.	Name	Ephemeroptera	Haplotaxida	Value
1	Paso de La Balsa	1	0	43
2	Paso de La Bolsa	1	0	42
3	Puente Hormiguero	1	0	43
4	Antes Navarro	1	0	29**
5	Juanchito	0**	1**	26**
6	Paso de La Torre	0	1	14
7	Mediacanoa	0	1	15**
8	Puente La Victoria	1	1	30**
9	Anacaro	1	1	27**

Intermediate situation scenario (Year 2015)

Sampling site		presence (1) or absence (0)		BMWP-Colombia
Nr.	Name	Ephemeroptera	Haplotaxida	Value
1	Paso de La Balsa	1	0	43
2	Paso de La Bolsa	1	0	42
3	Puente Hormiguero	1	0	43
4	Antes Navarro	1	0	30
5	Juanchito	1	0	28
6	Paso de La Torre	0	1	16*
7	Mediacanoa	0	1	20*
8	Puente La Victoria	1	0*	35*
9	Anacaro	1	1	30*

High investment scenario (Year 2015)

Sampling site		presence (1) or absence (0)		BMWP-Colombia
Nr.	Name	Ephemeroptera	Haplotaxida	Value
1	Paso de La Balsa	1	0	43
2	Paso de La Bolsa	1	0	42
3	Puente Hormiguero	1	0	43
4	Antes Navarro	1	0	30
5	Juanchito	1	0	29*
6	Paso de La Torre	0	1	20*
7	Mediacanoa	0	1	23*
8	Puente La Victoria	1	0*	36*
9	Anacaro	1	0*	31*

* Water quality improvement considering the current situation scenario

** Water quality deterioration considering the current situation scenario

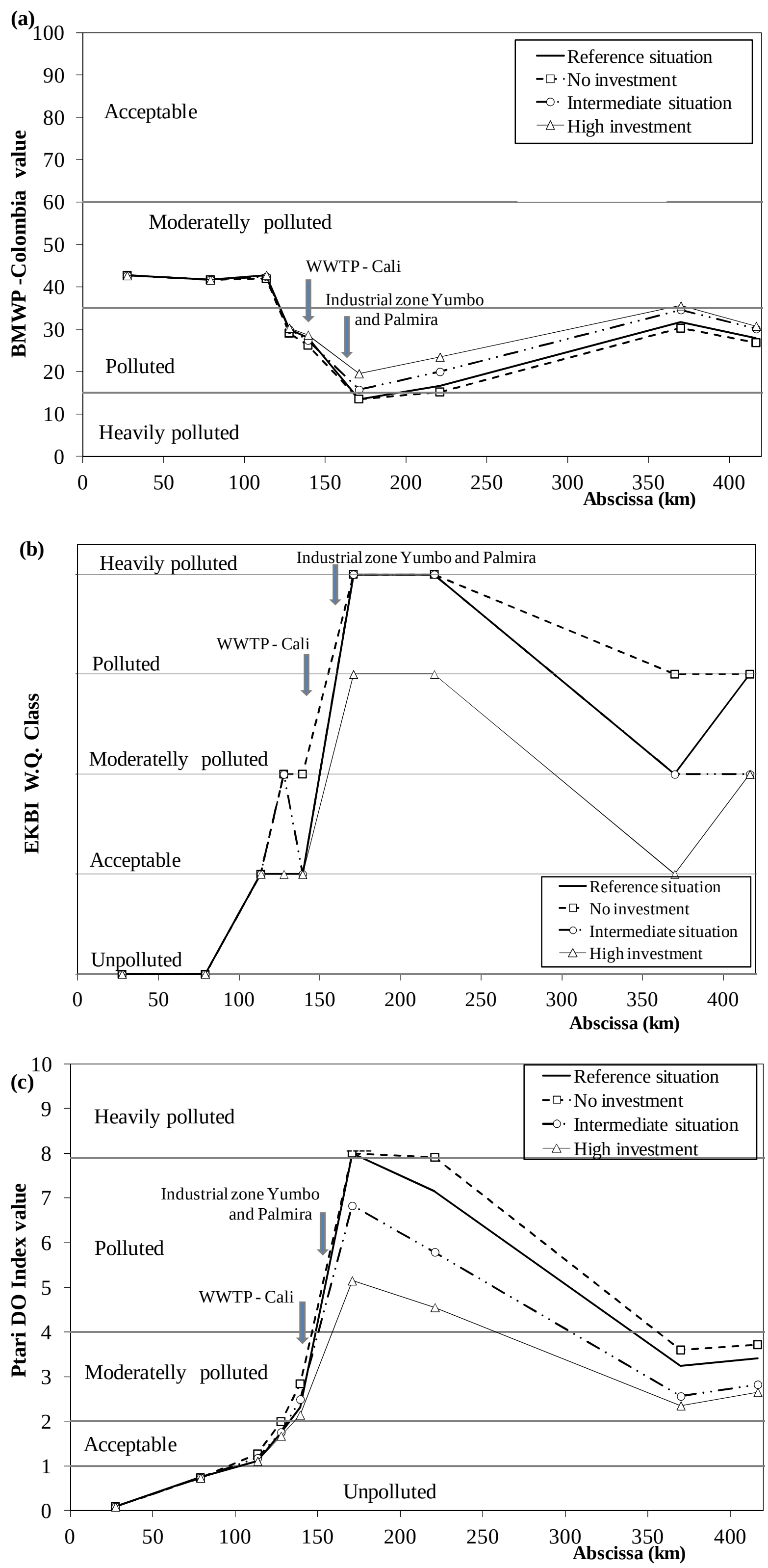


Fig. 3.4. Results of the application of the (a) BMWP-Colombia index predictive model, (b) the Oxygen Prati index and (c) the Expert knowledge based index (EKBI) for the scenarios considered in the Cauca river (see Table 3.2). The figure is divided in five zones going from unpolluted to heavily polluted The arrows indicate the presence of a WWTP (wastewater treatment plant) or an industrial zone.

3.4. Discussion

3.4.1 Habitat preference and ecological water quality

In the context of the Cauca river management, ecological assessments tools are needed to provide decision makers with accurate information about the EWQ, eventually to ensure habitat and species preservation. In order to manage conservation and to restore the river, it is necessary to find out the relationship between the water quality of the river (e.g. physicochemical and hydraulic variables) and the inhabiting. Models able to predict habitat requirements of organisms, may help to insure that planned actions reach the desired effects for the ecosystems (Ahmadi-Nedushan et al., 2006). The prediction of habitat conditions for two target macroinvertebrate taxa in the Cauca river, such as Ephemeroptera (pollution sensitive taxon) and Haplotaxida (pollution tolerant taxon), provide a good example of the applicability of ecological models.

The probability of occurrence of Ephemeroptera (i.e. Leptohyphidae and Leptophlebiidae families) in the Cauca river was mainly determined by DO and water depth, whereas for Haplotaxida this probability varied with DO (i.e. highest relative importance in the 95% confidence set of models; Table 3.4). The probability of occurrence of these two families of Ephemeroptera was positively related with DO and negatively related with water depth (Table 3.3), suggesting that these two families are more likely to be found in shallow sites of the Cauca river with high DO concentrations. Lock and Goethals (2011) stated that most of the species of Ephemeroptera, including those reported for the Leptohyphidae and Leptophlebiidae families, are only present at high DO concentrations and low conductivities. According to Dominguez et al. (2011a) and Lock and Goethals (in press), Ephemeroptera are characteristic for river sites with low human impact, having high DO and low BOD_5 concentrations. In the case of the Cauca river, high DO values indicate good physicochemical water quality, which supports that this taxon is pollution sensitive. It was also found that the presence of Haplotaxida was associated with low DO concentrations (Table 3.3), suggesting that this species can be present at river sites with high human impact, supporting the concept for this taxon as pollution tolerant. Regarding the BMWP-Colombia, the most important predictors were DO and water depth (Table 3.4). This index was positively related with DO and water velocity and negatively with water depth (Table 3.3). Similar results were reported by Dominguez et al. (2011b) who applied the same index in rivers in Ecuador, and reported that the index scored higher with increasing DO concentrations and high water velocities.

The water quality assessment of the Cauca river showed that the physicochemical indices over-predicted the water quality classes compared to the biological index BMWP-Colombia (Fig. C.2 in Appendix C). This over-prediction can be expected, because the ecological assessment provides more information on the state of an ecosystem than a physicochemical assessment alone. According to De Pauw and Hawkes (1993) the biotic component of an aquatic ecosystem can be considered as the “memory” of an ecosystem, integrating a wide range of ecological effects over time, while chemical analyses only provide information on the chemical water composition at the moment of sampling.

3.4.2 Model performance, uncertainty and validation

The predictions of occurrence of Ephemeroptera and Haplotaxida were determined accurately since the CCI, K and AUC for the averaged models met the criteria for a good model performance (CCI>0.7, K>0.4 and AUC>0.7, see Table 3.3). The predictions of the BMWP-Colombia index were less accurate, with 48 % of the variance (R^2) in the data being explained by the averaged model, mostly due to the variability that is inherently related to ecological data (Møller and Jennions, 2002; Symonds and Moussalli, 2011). Ecological models are simplified representations of the reality, thus, they can never fully predict nature and always contain errors in assumption, formulation and parameterization (Lek, 2007; Warmink et al., 2010). Therefore, uncertainty assessment of model simulations is important when models are used to support water management decisions (Beven and Binley, 1992; Refsgaard et al., 2007).

In general, the results of the calibration and verification process of the MIKE 11 model, showed that the model was able to accurately predict the dynamic tendencies and the maximum and minimum values of DO, BOD_5, temperature, flow, water depth and water velocity for the sampling stations of the Cauca river (see Figure 3.2). The uncertainty assessment based on the GLUE technique, showed that the model results were mainly in the range of the 95 % confidence bands, which indicates a good prediction capacity of the model. These bands allowed quantifying the reliability of the predictions and represent the influence of the uncertainty related with the values of the calibration parameters in each monitoring station of the river.

In this research the multi-model inference method based on the I-T approach (Burnham and Anderson, 2002) was used as equivalent to the multiple model simulation described by Refsgaard et al. (2007). This method allows selecting a set of “best models” (using the AICc and the goodness of fit) considering selection uncertainty. Specific percentages of the confident set of models for the “best models”; 95% and 85% for the LRM for Ephemeroptera and Haplotaxida respectively and 80% for the NBRM were estimated (see Table 3.3). As such, it is possible to be 95%, 85% and 80% confident that one of the models within this credibility set is the best approximating model. Additionally, full multimodel inference was estimated, such as full-model averaged predictions, considering the 95% of confident set of models. Model-averaged predictions are useful in contexts such as the one presented here, where there is reasonably high model uncertainty (i.e. the best AIC model is not strongly weighted), because predictions are not conditional on a single model (Burnham and Anderson, 2002). Model averaging recognises that there are two forms of uncertainty in modelling, the parameter uncertainty and the model uncertainty. The uncertainty in parameter estimates is measured by standard errors and confidence intervals for parameters. Model uncertainty considers that usually the ‘true’ model is unknown, and there is a probability that each candidate model is the ‘true’ model (Freckleton, 2011). When model uncertainty is present the I-T approach has considerable advantages over more traditional stepwise and null-hypothesis approaches to model selection, where we only end up with a single best model. Model averaged predictions are likely to be more robust than those derived from a single best model (Zuur et al., 2009). Moreover, keeping all the models from the best set of models, allowed picking a specific model with specific predictor variables based on considerations other than the statistical one, such as the ecological relevance of the predictors or the model applicability.

The MIKE 11 model was validated with an independent dataset and allowed evaluating the capacity of the calibrated model and predicting water quality under different hydraulic conditions from those used for the calibration. The validation of the ecological models (LRM and NBRM) was performed using two criteria: (1) a post-hoc evaluation of the model adequacy (Zuur et al., 2009; Fox and Weisberg, 2011) and; (2) evaluation of the predictive performance of the selected models (Gibson et al., 2004). For the first criteria, no patterns in the residual plots were found in the fitted smooth curve, which means that the LRM and NBRM are suitable for modelling the dataset (see Appendix C4-C6). For the second criteria, the selected models showed good model performances for the LRM and a moderate performance for the NBRM (see Table 3.3). However, the ecological models presented can still be improved in some aspects. Ideally the prediction capability of the models and the model averages would have been compared using an independent dataset. There is a general trend in the majority of ecological modelling studies to carry out model validation with independent data (Gibson et al., 2004). This was not possible in this study due to the limited dataset available. Therefore, the collection of an independent dataset in future studies will allow a full assessment of the adequacy of the ecological models. Changes in data collection strategy towards datasets where all variables (i.e. physicochemical, hydraulic and biological) are gathered during one sampling event are required.

3.4.3. Implementation of pollution control scenarios

Considering that the optimal balance between the different stakeholder activities needs an in depth insight in the integrated water resources management (Molle, 2009), it is vital that stakeholders participate in the modelling process (Voinov and Bousquet, 2010). Therefore, in this research four different scenarios for pollution control in the Cauca river basin were proposed by environmental authorities, municipalities and industries. In general, the scenarios showed that in spite of the reduction of the pollution load, the DO concentrations in the station Paso de La Torre (abscissa 170.8 km) for all proposed scenarios never reached values for DO higher than 2.6 mg/L (Fig. 3.3). Additionally, these DO values are still lower than the minimum standard value established by the Colombian legislation (i.e. Decree 1594 of 1984) for different uses of the water resource, which means, lower than 70% of the DO saturation concentration (5.2 mg/L O_2 for this river). The stretch located between the station Paso de La Torre and Mediacanoa (abscissa 220.9 km) is the most critical in terms of pollution, mainly because of the discharge of wastewater coming from the cities of Cali, Yumbo and Palmira. The habitat suitability models in these scenarios clearly indicated an improvement in potential habitat availability for the Ephemeroptera and a decrease in potential habitat for the Haplotaxida as the pollution load from domestic and industrial wastewaters is reduced.

The analysis of the water quality management scenarios presented in this study mainly dealt with physicochemical pollution. However, an improved data collection strategy will result in more consistent and larger datasets, allowing to consider also other types of pollution control such as the simultaneous effect of reducing the physicochemical pollution and enhancing the dilution capacity by increasing the minimum instream flow of the Cauca river (after the Salvajina dam).

3.4.4 Evaluation of the integrated ecological modelling framework

Nowadays, river quality assessment in Colombia relies mainly on physicochemical standards, however, there is a gap concerning the impact of different pressures on river biota, which are used to assess river water quality. Some of these pressures are physicochemical pollution, physical changes and anthropogenic manipulation of the aquatic habitat. The availability and use of decision support tools for water management, such as the one presented in this study, give an assessment of the impact of these pressures on river biota. By providing an integrated ecological modelling approach, the integration of different models, data and information resources is encouraged. This integrated approach serves, besides its function as a decision support tool, as a communication tool for providing information to the river managers.

In this research, the modular approach for model integration was implemented. This approach included an existing model for the hydrodynamic and physicochemical components (MIKE 11; CVC and Univalle, 2007) and new models (i.e. LRM and NBRM) for the ecological components were developed. This flexible integrated modelling framework allows updating or replacing these regression models by better models when available, without having to change the framework.

In the model development phase different combinations of physicochemical (i.e. DO) and hydromorphological variables (i.e. water depth and water velocity) were considered (see Table 3.3). These variables had low correlation among them and they were kept after the collinearity analysis. However, there are impacts such as nutrients (i.e. nitrogen and phosphorous), conductivity, particulate inorganic and organic matter, type of bank structure, type of substrate, water body slope and water body sinuosity that may influence the ecological state of rivers (Everaert et al., 2013). Therefore, in order to have a broad spectrum of the EWQ and to be able to construct more reliable models, more data should be collected in surface waters characterized by a very good or good EWQ and more physicochemical and hydromorphological variables need to be monitored. Thus, the MIKE 11 model could be used to simulate other processes and to predict some additional variables so that these can be included in the ecological models.

3.5 Conclusions

In this study, the IEMF proposed that integrates a hydraulic and physicochemical water quality model with aquatic ecological models was implemented and tested. The application of the IEMF in the Cauca river (Colombia) showed that the currently foreseen investments in sanitation infrastructure will lead to modest improvements of the EWQ. Therefore, further actions should be considered to achieve a good EWQ.

Chapter 4: Case study 2: Integrated ecological modelling for decision support in the water management of the Cuenca river in Ecuador

Adapted from:

Holguin-Gonzalez, J.E., Boets, P., Alvarado, A., Cisneros, F., Carrasco, M.C., Wyseure G., Nopens, I., Goethals, P.L.M. (2013). 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.

Chapter 4: Case study 2: Integrated ecological modelling for decision support in the water management of the Cuenca river in Ecuador

Abstract:

During the present study the IEMF presented in Chapter 1, was tested and validated in a shallow mountain river in a tropical region, the river Cuenca in the Andes of Ecuador. Two types of ecological models were developed, habitat suitability models to predict the occurrence of macroinvertebrates and ecological assessment models to predict the Biotic Integrity Index using aquatic invertebrates (IBIAP). Three wastewater management scenarios were tested. The different scenarios indicated that the foreseen investments in sanitation infrastructure will lead to modest improvements of the ecological water quality. This improvement (i.e. increase of the biotic index) was only identified in 6 of the 21 monitoring stations considered in the River Cuenca and its tributaries. Therefore, it is necessary to control the impact of the industrial wastewater discharges and the diffuse pollution at the upper catchment of the tributaries to achieve a good ecological state. It was found that species distribution models that predict the habitat suitability for selected species of macroinvertebrates, improved the understanding of the causal mechanisms and processes that affect the ecological water quality and shape macroinvertebrate communities in rivers. Simulations of pollution control scenarios implemented in the IEMF indicated an improvement in potential habitat availability for Trichoptera (pollution sensitive taxon) and a decrease in potential habitat for Physidae (pollution tolerant taxon) as the pollution load from domestic and industrial wastewaters is reduced.

4.1 Introduction

Water quality modelling is an effective tool to investigate and describe the ecological state of a river system and allows predicting changes in this state when certain boundary or initial conditions are altered. In order to manage conservation and restoration of a river, based on a good model representation, it is necessary to determine the relationship between the environmental conditions (e.g. physicochemical and hydromorphological conditions) and the occurrence of organisms inhabiting that river. Nevertheless, to date, few examples of the integration of hydromorphological, physicochemical and ecological models for decision support in river management have been reported. Authors have been focusing on two approaches: (1) either linking hydraulic models with habitat suitability indices (HSI) based on hydraulic habitat reference curves (e.g. water depth, water velocity and type of substrate) (e.g. Bockelmann et al., 2004; Tomsic et al., 2007) or; (2) using existing software (i.e. monolithic approach for model integration) such as the Water Framework Directive Explorer (WFD-Explorer) (Deltares, 2009). However, these approaches have limitations. On one hand, the HSI approach does not allow assessing simultaneously the impact of physicochemical pollution and hydromorphological disturbances on the habitat of aquatic species. On the other hand, the WFD-Explorer considers the impact of these two river pressures, but it operates on a coarse river basin scale level, whereas the impact of physical habitat changes on river biology occurs at smaller scale level (Mouton et al., 2009a). Additionally, the WFD-Explorer simplifies water quality processes as a retention factor.

The limitations of these two approaches emphasize the need for the development of a detailed physical habitat and water quality model that allows assessing simultaneously the impact of hydromorphological pressures and physicochemical pollution on the ecological water quality of a river. This study describes the implementation and validation of the IEMF presented in Chapter 1, applied on the River Cuenca, an Andean mountain river (average altitude of 2.550 meters above sea level) in Ecuador. In this study, three of the four basic modelling components of the IEMF (see Fig. 1.1, in Chapter 1) were considered. The first and second components, which correspond to river water quantity and quality modelling, were included in the QUAL2Kw model (Pelletier et al., 2006). The third component included two types of ecological models based on data-driven modelling techniques. The first ecological model allowed predicting the presence of two target macroinvertebrate taxa (i.e. Trichoptera and Physidae) based on logistic regression. The second model allowed predicting the Biotic Integrity Index using aquatic invertebrates – IBIAP (Carrasco, 2008) based on model trees. The impact of different wastewater treatment/disposal strategies on the ecological state of the receiving river was evaluated.

4.2 Materials and methods

4.2.1 Study area

The River Cuenca is an Andean mountain river formed by the confluence of four rivers: the Tomebamba, Tarqui, Yanuncay and Machangara rivers. These cross the city of Cuenca in the southern Province of Azuay in Ecuador (Fig. 4.1) and come together in the lower part of the city. Cuenca is the third largest city in Ecuador with around 400,000 inhabitants (Carrasco, 2008) and the main urbanization centre in the study area. This study focuses on the river network with a total length of 63.5 km around this city in a basin area of about 1500 km^2. The elevation of the sampling sites at the study zone varies from 2750 to 2318 meters above sea level. Water is extracted for drinking water, and to a lesser extent for industrial and agricultural water supply from the rivers Tomebamba and Yanuncay downstream of a nature reserve, called Cajas, which is located upstream of the city.

In the study area, the Tarqui river shows evidence of high organic pollution caused by uncontrolled diffuse fluxes from extensive livestock in the middle part of the catchment (rural area) and urban discharges. The other rivers have in their upstream (less populated) part a better water quality (Carrasco, 2008). Despite the sewer system for the collection of wastewater in Cuenca, there are still a number of diffuse and point sources of pollution from some Cuenca city districts that are affecting the water quality of these rivers. The Machangara river’s flow regime is highly influenced by two hydropower dams located 30 km upstream of Cuenca.

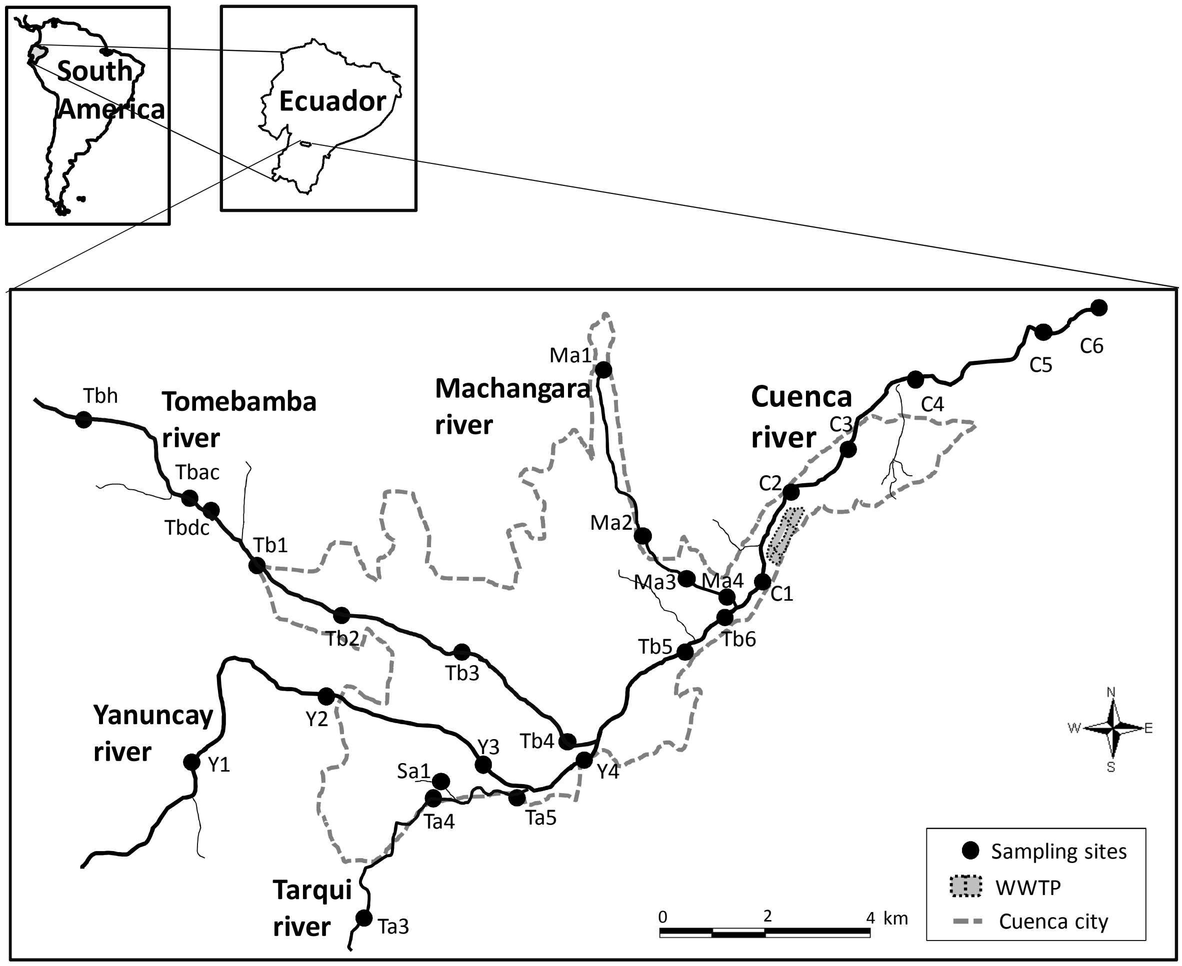


Fig. 4.1. Overview of the study area and monitoring stations in the Cuenca river basin in the Azuay Province, Ecuador (WWTP: wastewater treatment plant, Ta3-Ta5: Tarqui river, Y1-Y3: Yanuncay river, Ma1-Ma4: Machangara river, Tb1-Tb6: Tomebamba river, C1-C6: Cuenca river).

The sanitation company in the city of Cuenca (ETAPA) has been operating since the year 1984 to improve the water quality of the four rivers which cross the city. This company has been investing in infrastructures used for environmental protection, such as the collection and treatment of wastewater (ETAPA, 2007). Cuenca has a combined sewerage flowing into a waste stabilization pond system, which has been in operation since 1999 (ETAPA, 2009). The system comprises of aerated, facultative and maturation ponds in series and there is a discharge of around two tons of organic matter per day in terms of five-day biological oxygen demand (BOD_5) to the River Cuenca (ETAPA, 2009). Moreover, part of the wastewater from the sewer system is directly discharged into the River Cuenca or its tributaries without any treatment. Nitrogen and phosphorus balances have been altered by agricultural run-offs and urban sewage discharges. These discharges of combined treated and untreated wastewater cause an increasing deterioration of the water quality of the River Cuenca and can potentially affect human health and aquatic life, limit water uses, affect river ecology and cause loss of amenity. Nowadays, ETAPA is interested in an integrated urban water system model of the River Cuenca for a cost-efficient wastewater treatment optimization which respects the ecological aspects. This should avoid this pollution problem to become critical in the near future, especially during the dry season (low flow rates in the river).

4.2.2 Data collection, coupling of data and dataset pre-processing

The dataset used in this research was collected during 1997-2008 by ETAPA and by the authors during the year 2009. The study system consisted of 27 sites (Fig. 4.1) with long term monitoring data, however, only of 20 sampling locations biological information was available. The biological dataset comprised of 88 samples of macroinvertebrates. These samples were taken at the aforementioned 20 sites, all of which were assessed at least once during this period. In this dataset some physicochemical or hydraulic variables were not measured for one or more biological samples (incomplete measurement campaign). As a consequence, the data in this study was limited to the records that contained information of all variables (complete measurement campaign). Thus, 60 macroinvertebrate samples were retained after coupling the physicochemical and hydraulic information to the biological data.

In order to enable the coupling of the ecological models developed with the water quality model QUAL2Kw, a dataset containing hydraulic and physicochemical data was built exclusively including variables modelled by the QUAL2Kw model (i.e. dissolved oxygen (DO), temperature, BOD_5, Faecal Coliforms (FC), flow, water depth and water velocity) and the biological data. The final dataset consisted of seven predictor variables (four physicochemical and three hydraulic variables) and three ecological response variables (see Appendix B; Table B.2). The latter were: the value for the biotic index and the presence/absence of two different target taxa of macroinvertebrates. These two taxa of macroinvertebrates were: Trichoptera (pollution sensitive taxon), which is a biological indicator for good water quality conditions and Physidae (pollution tolerant taxon), which is a biological indicator for polluted water with a high organic matter content (Carrasco, 2008). The two selected macroinvertebrate taxa are complementary biological indicators, because their geographic distribution in the River Cuenca and its tributaries (presence or absence) depends on their pollution tolerance. The pollution tolerance scores (PTS) ranges from ten for very pollution sensitive to one for very pollution tolerant taxa. According to the biotic index IBIAP (Carrasco, 2008), the respective PTS for the Trichoptera families identified in this river (Polycentropodidae, Limnephilidae, Leptoceridae, Hydrobiosidae, Hydroptilidae, Philopotamidae and Calamoceratidae) lies between seven and ten, whereas for Physidae the PTS is three.

Concerning pre-processing of the data used to build the ecological models, it was focussed on three aspects: (1) evaluation of possible outliers, (2) evaluation of the collinearity and (3) relationships between the response variable and the predictor variables. The evaluation of outliers was performed using two graphical tools, box plots and Cleveland dot plots (Zuur et al, 2010). In order to avoid high collinearity between the predictor variables, a procedure based on a Principal Component Analysis (PCA) with a varimax rotation and the non-parametric correlation coefficient Kendall's (τ) were used. The varimax rotation in the PCA allowed maximising the independence of the Principal Components (PCs). To explore the correlation between the potential predictor variables used to build the models, this coefficient (τ) was used rather than the Pearson correlation coefficient, because the first can deal better with outliers and extreme distributions of the variables (Willems et al., 2008). Based on the PCA and the correlation analysis different sets of predictor variables were offered to the selection algorithms of the ecological models (see Appendix D and section 4.3.1).

4.2.3 Water quality modelling techniques used

The three modelling components of the IEMF considered for this study were: (1) a river water quantity model, (2) a river water quality model and, (3) river habitat suitability and ecological assessment models. For the first and second components, the hydraulic and physicochemical water quality model QUAL2Kw (Pelletier et al., 2006) was used. For the third component, logistic regression models (LRM) for presence/absence predictions and model trees for IBIAP index predictions, were implemented. Once the integration of models is performed, they can be used for simulations of scenarios for water management plans. The ecological models developed were applied on the resulting hydraulic and physicochemical data of the QUAL2Kw model. Different datasets were generated based on the outcome of the different wastewater treatment scenarios.

4.2.3.1 Hydraulic and physicochemical water quality model

In order to perform the water quality and hydraulic modelling in the Cuenca rivers, the QUAL2Kw (Pelletier et al., 2006) model was implemented. QUAL2Kw is an adaptation from the QUAL2K (Chapra and Pelletier, 2003) and a modernized version of the QUAL2E (Brown and Barnwell, 1987). QUAL2E is a standard river water-quality model developed by the United States Environmental Protection Agency (US EPA). Chapra and Pelletier (2003) developed the QUAL2K with several new features compared with QUAL2E that allow them to be applied to shallow, upland streams. The QUAL2K includes several enhancements: more flexible model segmentation, the simulation of two types of carbonaceous biological oxygen demand (CBOD), fast CBOD (CBODf) and slow CBOD (CBODs), oxygen attenuation of oxidation reactions and simulation of sediment fluxes, bottom algae, pH, and a generic pathogen indicator for bacteria (Chapra and Pelletier, 2003; Pelletier and Chapra, 2005). QUAL2Kw added to QUAL2K two new major features: the option of simulation of dendritic water systems and the inclusion of an autocalibration routine based on a genetic algorithm.

QUAL2Kw uses a general equation of mass balance for the concentration of a constituent c_i in the water column (excluding hyporheic exchange) in a reach i (Pelletier et al., 2006), which is written as:

$$\frac{dc_i}{dt} = \frac{Q_{i-1}}{V_i} c_{i-1} - \frac{Q_i}{V_i} c_i - \frac{Q_{ab,i}}{V_i} c_i + \frac{E'_{i-1}}{V_i}\left(c_{i-1} - c_i\right) + \frac{E'_i}{V_i}\left(c_{i+1} - c_i\right) + \frac{W_i}{V_i} + S_i \qquad (4.1)$$

where Q_i is the flow [m^3/d, *ab* is the abstraction], V_i is the volume (m^3), E'_i is the bulk dispersion coefficient between reaches i and i +1 [m^3/d], W_i is the external loading of the constituent to reach i [g/d or mg/d], and S_i are sources and sinks of the constituent due to reactions and mass transfer mechanisms [g/m^3/d or mg/m^3].

QUAL2Kw is a one-dimensional and steady state flow water quality model for streams and rivers, programmed in Visual Basic for Applications (VBA). The software Microsoft Excel is used as the graphical user interface for input, running the model, and presenting the output. The numerical integration (i.e. Euler, fourth-order Runge-Kutta, or adaptive method) is performed by a compiled Fortran 95 program that is run by the Excel VBA program (Pelletier et al., 2006). The spatial approximation of only 1 dimension (1D) is considered appropriate since the river-reaches are long relative to the mixing length over the cross-section and the transport of contaminants is dominated by longitudinal liquid motion.

The QUAL2Kw model represents a river as a series of reaches. These represent stretches of a river that have constant hydraulic characteristics (e.g., slope, bottom width, etc.). The model simulates dendritic water systems, i.e. those where simulation extends not only to the main stream, but also to its tributaries. The model is capable of simulating one (1) main stem and three (3) tributary streams. Tributaries can be operated independently or integrated into the main branch depending on user needs. In this research, the confluence of the Tomebamba and Cuenca rivers was considered as the main stem (length of 27.5 km). The three tributary rivers called Yanuncay (length of 9.5 km), Tarqui (length of 15.5 km) and Machangara (length of 11.0 km) were modelled individually and integrated into the main branch. The result of the last computational element of the tributary river was seen as an input (i.e. point source) for the main stream. The length of the rivers was divided into 19 (Tarqui river), 31 (Yanuncay river), 22 (Machangara river) and 55 (confluence of the Tomebamba river, element 1 to 32 and Cuenca river, element 33 to 55) sub reaches with a length equal to 0.5 km each. In total, a river length of 63.5 km was modelled. Fig. 4.2 shows the segmentation and position of the main discharges along these rivers.

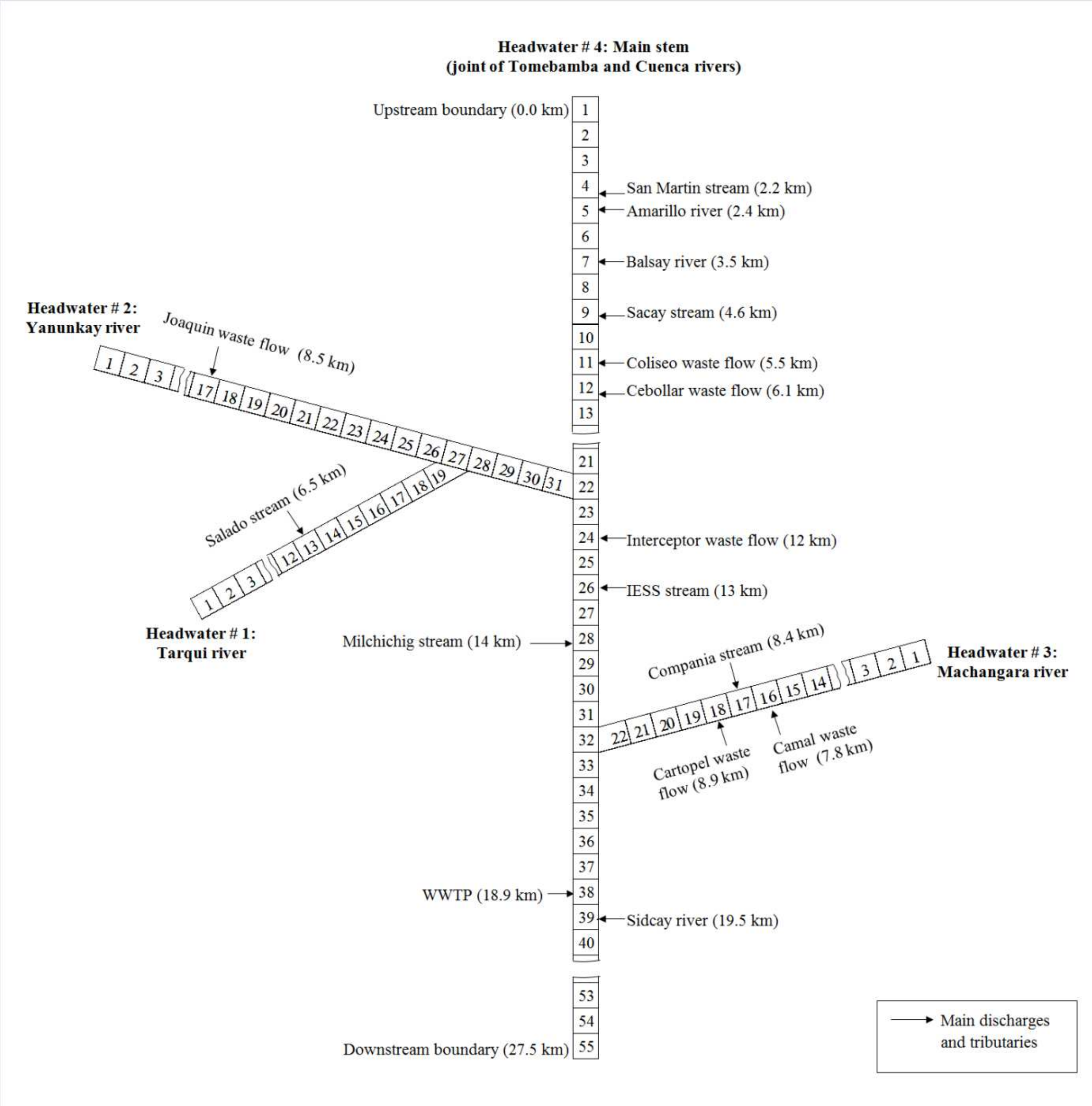


Fig. 4.2. System segmentation with location of the main pollution sources along the Tarqui, Yanuncay, Machangara, Tomebamba and Cuenca rivers. The main stem is the joint of the Tomebamba (elements 1 to 32) and Cuenca (elements 33 to 55) rivers (WWTP: wastewater treatment plant).

For the calibration and validation of the QUAL2Kw the data collected by ETAPA during nine intensive monitoring campaigns from July to October 2001 in the River Cuenca and its tributaries were used. These monitoring campaigns included the measurement of DO, temperature, BOD_5, FC, flow, water depth and water velocity. They considered the most important input pollution loads and based on those, samples were taken sequentially accounting for the average water velocity along the river. Thus, calibration and validation monitoring campaigns were performed following a water-mass volume from the source to the mouth. This is a key aspect for implementing steady state models such as the QUAL2Kw, in which flow and water quality variables are assumed constant in time (Díaz-Granados et al., 2009). For the calibration dataset of the QUAL2Kw, the average values of the physicochemical and hydraulic variables measured during four of those nine monitoring campaigns were taken. These four campaigns had similar input pollution loads and were monitored during dry flow conditions. Validation of the model was performed using the five remaining campaigns, which both included dry season conditions as well as wet season conditions. The variables modelled by the QUAL2Kw were DO, temperature, BOD_5, FC, flow, water depth and water velocity.

Calibration of the physicochemical variables in the QUAL2Kw model was performed by a constraint-based random search method (Oddi et al., 2005) (see section 3.2.4.1 in Chapter 3. for details of this method). The calibration ranges were estimated for each kinetic rate parameter considering the minimum and maximum values reported by Pelletier and Chapra (2005); Kannel et al. (2007) and Cho and Ha (2010). The calibration parameters (kinetic rates) considered were: settling velocity, CBODf oxidation rate, CBODs hydrolysis rate, CBODs oxidation rate, pathogens decay rate, pathogens settling velocity and alpha constant for light mortality of pathogens. To calculate the re-aeration rate, the Owens–Gibbs formula (Owens et al., 1964) was applied in the Tarqui and Machangara rivers, whereas the Churchill formula (Churchill et al., 1962) was used in the Yanuncay, Tomebamba and Cuenca rivers. These equations were selected according to the range of depths and velocities encountered in the rivers (Chapra, 1997). The Owens–Gibbs formula is appropriate for shallow streams (0.12-0.73 m) with slow velocities (0.03-0.55 m/s), whereas the Churchill formula is appropriate for deeper streams (0.61-3.35 m) and higher velocities (0.55-1.52 m/s). The other rate parameters were retained at their default values in the QUAL2Kw model. To calibrate the hydraulic characteristics of the QUAL2Kw, the depth-discharge and velocity-discharge curves that were derived using flow measurements and river geometry were used (ETAPA, 2007). For the validation process the model was run using the data of the five monitoring campaigns mentioned before, without changing the calibrated parameters. Additionally, uncertainty analysis was performed using the GLUE method (Beven and Binley, 1992), based on the results of the constraint-based random search method.

The evaluation criteria considered during the calibration and validation of the QUAL2Kw were the determination coefficient (R^2) and the modified index of agreement (*dm*). The R^2 and *dm* were calculated for each of the thousand simulations performed during the calibration and for each modelled variable. A weighted sum of the evaluation criteria was calculated for the three modelled variables (DO, BOD_5 and FC) and the model with the highest R^2 and *dm* for the three variables simultaneously was selected, leading to the best combination of values of the most sensitive calibration parameters. The R^2 is a measure of the goodness of fit of the regression model and is defined as the squared value of the coefficient of correlation according to Bravais-Pearson (Krause et al., 2005). *dm* is a dimensionless indicator widely used to evaluate the goodness-of-fit of hydrologic and water quality models (Krause et al., 2005; Harmel and Smith, 2007). This index is a modified version of the index of agreement (*d*; Willmott, 1981) that uses the absolute value of the deviations instead of the squared deviations (Legates and McCabe, 1999; Harmel and Smith, 2007). R^2 and *dm* range between zero and one, and the closer the value to one, the better the model predicts the training (calibration) or validation data.

4.2.3.2 Ecological modelling

Macroinvertebrate predictive models: LRM allow predicting the probability of a species occurrence or distribution (Rushton et al., 2004; Ahmadi-Nedushan et al., 2006). This data-driven method is easy to use for the analysis of dichotomous (presence/absence) data and is implemented in many software packages. Details about the implementation of the LRM are presented in Appendix A. All hydraulic and physicochemical variables were considered for inclusion in the LRM through a stepwise variable selection process with statistical considerations, in a multivariable logistic regression analysis, implemented in the statistical software XLSTAT version 2010 (Addinsoft, 2010). The criterion for removal of variables was based on statistical considerations using the likelihood ratio test with a significance level of $p > 0.05$. Models were fitted using the maximization of the likelihood function (McCullogh and Nelder, 1989) using the Newton-Raphson algorithm.

To test the robustness of the models, the LRM constructed were validated based on a three-fold cross validation. The total dataset was, after reshuffling, split in three subsets: two thirds were used for training and one third for validation. For each training and validation set a model was built and in this way, a performance value for each of the three different models was obtained. The results from the three-folds were averaged to produce a single prediction of the dependent variable. If the predictive performance of the model for each fold was similar, a final model was constructed with all the data.

To assess the model performance of the LRM three criteria calculated from the confusion matrix were evaluated: (1) the percentage of Correctly Classified Instances (CCI); (2) Cohen's kappa coefficient (Cohen's *K*) and (3) the area under the receiver-operating-characteristic (ROC) curve called AUC. Details about the ranks of model performance considering Cohen's *K*, ROC and AUC values are described in Chapter 3 and Appendix A.

Predictive model for the biotic index: The second data-driven modelling technique implemented (i.e. model trees) allows performing a biological assessment by predicting the value of the biotic index IBIAP (Carrasco, 2008), based on abiotic river conditions (physicochemical and hydraulic variables). The IBIAP index uses the following environmental response variables: the species richness, the number of EPT taxa (Ephemeroptera, Plecoptera and Trichoptera), the number of filterers and shredders and the mean pollution tolerance of the sample. The final result of the IBIAP index is an integer value between 0 and 16. A high ecological water quality has a IBIAP value of 16, a good quality has a value between 12 and 15, a moderate quality has a value between 6 and 12 and a poor water quality has a value lower than 6.

Decision tree learning is one of the most popular machine learning techniques used in ecological modelling (Debeljak and Džeroski, 2011). Decision trees are mostly used for predictive modelling and for extracting new knowledge about the observed processes. The basic idea of generating decision tree models is to develop simple and transparent models that are easy to use and interpret. These models are generated through an iterative splitting of data into subspaces of the whole attribute space, where the goal is to maximize the distance between groups at each split (Stravs et al., 2008). Decision tree models, allow representing a series of rules that lead to a class value, numerical value or linear equation, and are therefore classified into: classification trees (CT) with class values as leafs of the model; regression trees (RT) with constant numerical values as leafs of the model and; model trees (MT) with linear equations as leafs of the model (Stravs et al., 2008).

In a preliminary assessment (i.e. without taking into account three-fold cross-validation: see details further) the applicability of the three types of decision tree models with the dataset was evaluated and MT gave the best results. Therefore, MT with the classifier algorithm M5P implemented in the Waikato Environment for Knowledge Analysis (WEKA) (Witten et al., 2011) were used. The minimum number of instances to allow at a leaf node was fixed as four and an unpruned tree with unsmoothed predictions was generated. The MT allowed predicting the value of the biotic index IBIAP based on physicochemical and hydraulic variables. The linear models in the leaves of the MT explain the response variable Y (i.e. IBIAP value) by a vector of n predictor variables $X = X_1, X_2, ..., X_n$ (e.g. DO, BOD_5, water velocity). The MT were trained and evaluated based on a three-fold cross validation procedure. The performances of the MT were assessed by the Pearson correlation coefficient (r) and R^2 for the predictions of the IBIAP values and by the CCI for the predictions of the IBIAP classes.

Over the last decade, applications of machine learning techniques such as CT, RT and MT in an ecological context have been reported by several authors. De'ath (2002) described the relationships between environmental characteristics and species by means of RT. Pesch and Schröder (2006) used this approach to relate the risk of metal bioaccumulation with site-specific and eco-regional characteristics. Stravs et al. (2008) presented an application of CT and RT for the analysis of the process of precipitation interception by a forest in a river basin. Kocev et al. (2009) used RT to model the quality of vegetation based on GIS-data. Boets et al. (2010), Everaert et al. (2011) and Boets et al. (in press) implemented CT to analyze the impact of aquatic invasive species on the native communities. Debeljak and Džeroski (2011) presented a review of the applications of different types of decision trees (i.e. CT, RT and MT) in ecological modelling. These applications include modelling population dynamics and habitat suitability for different organisms in different ecosystems exposed to different environmental pressures. There are several studies comparing different multivariate statistics and machine learning techniques (Vayssières et al., 2000; Guisan et al., 2007; Meynard and Quinn, 2007; Pearson et al., 2006; Segurado and Araujo, 2004). The choice of model type has much to do with availability of information and software, current fashion and, of course, with the specific aim of the study.

4.2.4 Simulation of pollution control scenarios

Once the logistic regression model and model trees are developed, they can be used to make predictions about the dependent variables (i.e. macroinvertebrate taxa presence/absence and IBIAP values) based on other independent values than the values that were used to build the models. Using the integrated ecological model, three scenarios were run and evaluated. In Scenario 1, the discharge of untreated wastewater was considered, this would be the case if the water would pass via the bypass system in the wastewater treatment plant (WWTP) or in case the WWTP did not function at all properly. In Scenario 2, the annual averaged functioning condition of the WWTP was simulated. A reference situation in the year 2009, with a removal efficiency of 84% in BOD_5 and two logarithmic units in the FC (ETAPA, 2009) was considered. In this scenario, 85% of the total amount of domestic wastewater (DWW) produced by the city is collected and treated in the WWTP (pers. com., A. Alvarado). In Scenario 3, the collection and treatment of the total amount of DWW was evaluated (100% of the wastewater is treated). The procedure followed for the simulation of these scenarios was to use the output data of the QUAL2Kw (water quality and quantity variables) as input data for the ecological models (LRM and MT). Since the QUAL2Kw is a steady flow stream water quality model, the resulting data (hydraulic and physicochemical data) was considered as daily average data for the LRM models and MT, in all sampling points of the system modelled.

4.3. Results

4.3.1 Data analysis and variable selection

The evaluation of possible outliers in the dataset with the 60 samples that contained physicochemical, hydraulic and biological information showed that there were no outliers. Regarding the PCA, the first six PCs explain 95% of the variance in the data and the variables flow and water velocity were included in the same PC. However, the correlation between these two variables ($\tau = 0.63$) was moderate ($0.45 \leq \tau \leq 0.7$), therefore, it was decided to keep both as predictor variables. The rest of the variables were not correlated ($-0.2 \leq \tau \leq 0.2$) or slightly correlated ($-0.45 \leq \tau \leq -0.2$ and $0.2 \leq \tau \leq 0.45$) with each other. BOD_5 showed the highest correlation with the IBIAP index ($\tau = 0.55$), whereas flow, water velocity and water depth showed the lowest correlation with this index (τ = 0.04, 0.16 and 0.2 respectively).

4.3.2 Hydraulic and physicochemical water quality model

The results of the calibration and verification processes of the QUAL2Kw, showed that the water quality model reproduces with good precision the tendencies and the maximum and minimum values of DO, BOD_5 and FC in the monitoring stations of the rivers situated in the city of Cuenca. As example, the results of the calibrated model for dry conditions (i.e. averaged values of the physicochemical and hydraulic variables measured during four monitoring campaigns) are presented in Fig. 4.3. This condition is the most critical in terms of the water quality and quantity of the streams. Similar graphs were built for the hydraulic variables, but are not shown in this document. Model performance indicators and Standard Deviation (SD) obtained during the calibration and validation processes for DO, BOD_5, FC and flow can be seen in Table 4.1. The assessment of the reliability of the QUAL2Kw model showed that in the calibration dataset the model performed very well, with *dm* in the range between 0.81 and 0.94 and R^2 between 0.87 to 0.98, whereas for the validation set the model performance was somewhat lower but still sufficient, with *dm* in the range between 0.71 and 0.98 and R^2 between 0.72 and 0.91.

4.3.3 Modelled habitat preference and ecological assessment model

Modelled habitat preference for the targeted macroinvertebrates: The prevalence for Trichoptera and Physidae taxa in this study sites was 38% and 70%, respectively. Trichoptera were mainly present in sampling locations where the ecological water quality was good to high (mainly IBIAP values higher than 8), whereas Physidae were present where the water quality was poor to moderate. After selecting the set of explanatory variables for the best logistic regression model, it was found that the most important variable that determined the presence of Trichoptera in the Cuenca river was BOD_5 (p-value < 0.0001), whereas for Physidae the most important variable was the number of Faecal Coliforms (p-value < 0.0001) (Fig. 4.4). The LRM for the two taxa selected were:

$$p_{(\text{LR, Trichopter a})} = \frac{1}{1 + \exp(-2.065 + 1.015\,BOD_5)} \tag{4.2}$$

$$p_{(\text{LR, Physidae})} = \frac{1}{1 + \exp(0.469 - 1.53 \times 10^{-5}\,\text{FC})} \tag{4.3}$$

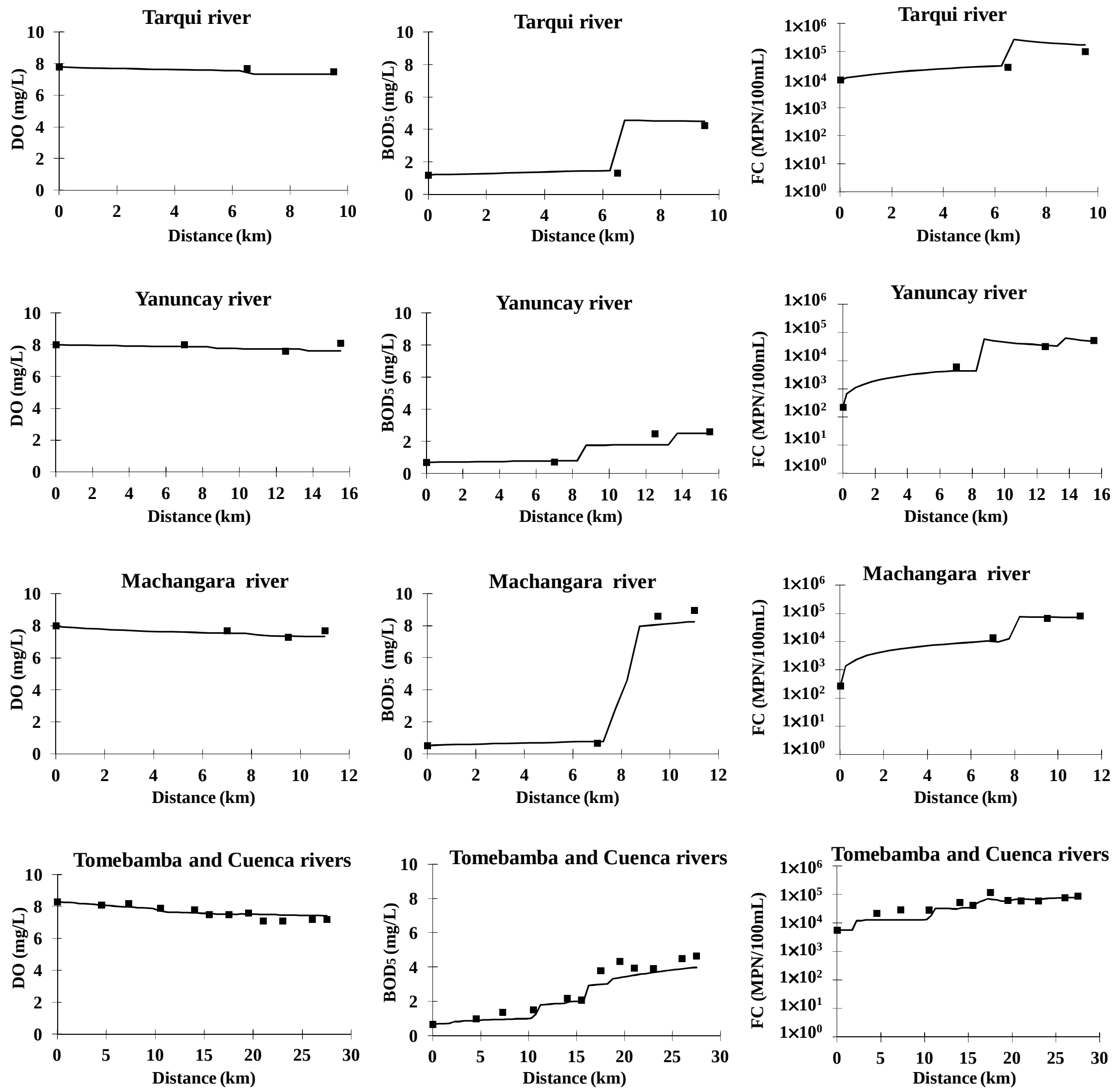


Fig. 4.3. Calibration results of the QUAL2Kw for the Tarqui, Yanuncay, Machangara and the joint of the Tomebamba and Cuenca rivers during dry season for dissolved oxygen (DO), five-day biological oxygen demand (BOD_5) and Faecal Coliforms (FC). Distance measured from the first station.

Table 4.1. Average model performance indicators and Standard Deviation for the water quality model QUAL2Kw in the calibration and validation dataset (*dm*: modified index of agreement; R^2: determination coefficient; BOD_5: five-day biological oxygen demand).

Variable	Criterion	Calibration	Validation
Dissolved oxygen	*dm*	0.81	0.71 ± 0.11
	R^2	0.87	0.75 ± 0.09
BOD_5	*dm*	0.91	0.79 ± 0.13
	R^2	0.98	0.72 ± 0.19
Faecal Coliforms	*dm*	0.91	0.84 ± 0.15
	R^2	0.97	0.77 ± 0.05
Flow	*dm*	0.94	0.98 ± 0.01
	R^2	0.97	0.91 ± 0.03

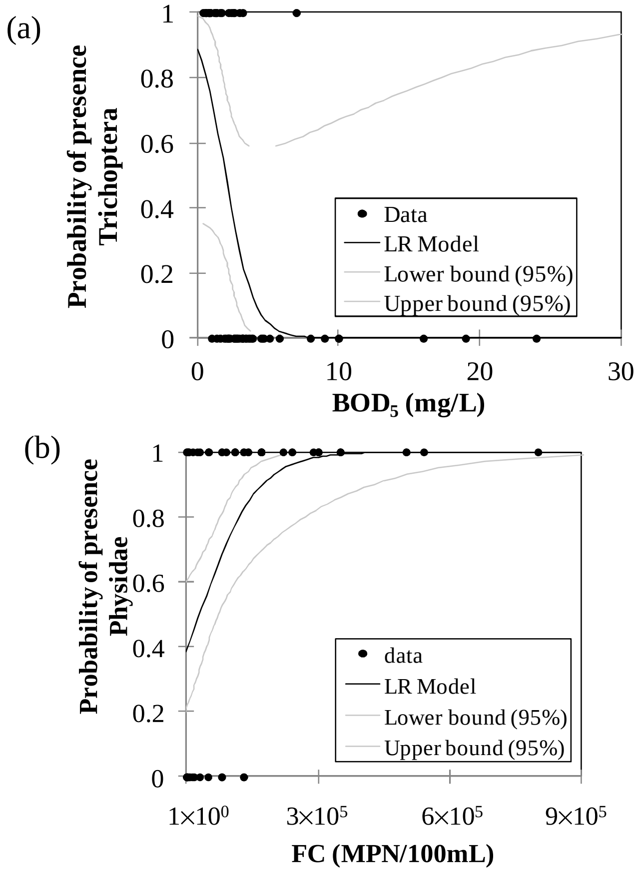


Fig. 4.4. Logistic regression model (LRM) predicting the presence or absence of Trichoptera (a) and Physidae (b) in the River Cuenca and its tributaries, according to equations 4.2 and 4.3.

The probability of occurrence of Trichoptera in the Cuenca river was negatively related with BOD_5 (Fig. 4.4a), suggesting that this taxon is more likely to be found in sampling sites with a low organic pollution level (low BOD_5 concentrations), supporting the concept for this taxon as pollution sensitive. It was also found that the presence of Physidae was associated with high levels of Faecal Coliforms (Fig. 4.4b), suggesting that this taxon can be present at river sites with high human impact, supporting the concept for this taxon as pollution tolerant. The 95% confidence interval for the LRM for Trichoptera and Physidae taxa indicates that there is more uncertainty in the LRM for Trichoptera (wider band), especially for BOD_5 values higher than 5 mg/L (upper bound of the 95% CI confidence interval). This high uncertainty could be related with the influence of one record with a BOD_5 value of 7 mg/L which reported the presence of Trichoptera, however as can be seen from Fig. 4.4a, the probability of presence of Trichoptera at BOD_5 values higher than of 5 mg/L is lower than 10%, suggesting a possible outlier. To validate the model performance, a set with data independent from the training set is required. This is called as 'test' set (sometimes also termed 'validation' data), whilst data used to build the model can be called 'training' set (sometimes termed 'calibration' data). The assessment of the reliability of the LRM models (Table 4.2, complete dataset) showed that the models for Physidae (CCI=75%, K=0.41 and AUC=0.82) and Trichoptera (CCI=80%, K=0.57 and AUC=0.87) have a reasonable discrimination capacity and correctly discriminate between occupied (presence) and unoccupied (absence) sites in the dataset.

Ecological assessment model: A MT was built to understand the relationship between the biological water quality (expressed as the IBIAP index) and the physicochemical and hydraulic variables modelled with QUAL2Kw. The assessment of the reliability of the MT showed that in the training dataset the model performed well, with r = 0.82 ± 0.03, R^2 = 0.68 ± 0.05 and CCI = 81% ± 0.03, whereas for the test set the model performance was somewhat lower, r = 0.47 ± 0.19, R^2 = 0.25 ± 0.16 and CCI = 61% ± 0.17. Nevertheless, the predictive performance of the MT constructed with all the data was good, with r = 0.89 and R^2 = 0.80. In 84% of the cases the predicted water quality class was the same as the one measured (i.e. CCI of 84%), therefore it was decided to use this final model. The outputs of the MT obtained in this study were three linear equations (Table 4.3) which included six variables (i.e. temperature, BOD_5, DO, flow, water depth and water velocity). The most important variable in the MT was BOD_5 and depending on the value of this variable one specific linear model should be used (Fig. 4.5). Thus, if BOD_5 is lower than

1.8 mg O_2/L the linear model 1 (LM1) should be selected, otherwise if BOD_5 is higher than 1.8 mg O_2/L but lower than 7.5 mg O_2/L the LM2 is chosen and finally, in case that BOD_5 is higher than 7.5 mg O_2/L the LM3 is used. The threshold BOD_5 value of 1.8 is close to the one proposed by Chapman (1996) of 2 mg O_2/L or less for waters with a low pollution level, which indicates the ecological relevance of the model.

Table 4.2. Average model performance indicators and Standard Deviation for the logistic regression models (LRM) in the training, test and complete dataset. (CCI: Correctly Classified Instances; *K*: Cohen's kappa coefficient; AUC: area under the receiver-operating-characteristic curve). Good model performances in LRM are represented by CCI>0.7, *K*>0.4 and AUC>0.7.

LRM	CCI (%)	*K*	AUC
Training set			
Physidae	75.8 ± 5.8	0.43 ± 0.11	0.82 ± 0.04
Trichoptera	80.0 ± 0.0	0.54 ± 0.03	0.85 ± 0.03
Test set			
Physidae	75.0 ± 13.2	0.39 ± 0.23	0.80 ± 0.13
Trichoptera	81.7 ± 7.6	0.62 ± 0.16	0.93 ± 0.05
Complete dataset			
Physidae	75.0	0.41	0.82
Trichoptera	80.0	0.57	0.87

Table 4.3. Detailed linear equations obtained based on the model tree (MT). Considering $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n$, each equation explains the response variable *Y* (IBIAP value) by a vector of predictor variables $X = X_1, X_2, ..., Xn$, (Temp.: Temperature; DO: dissolved oxygen; BOD_5: five-day biological oxygen demand; V: Velocity) and β_0 (as intercept) and $\beta = \{\beta_1, ...\beta_m\}$ as regression constants.

Rule	Linear model	Intercept	Temp.	DO	BOD_5	Flow	Depth	V
$BOD_5 \leq 1.8$ mg/L	LM1	9.861	-0.121	*	-2.061	-0.099	4.211	1.119
$1.8 < BOD_5 \leq 7.5$ mg/L	LM2	6.078	-0.067	0.109	-0.007	0.013	0.556	1.287
$BOD_5 > 7.5$ mg/L	LM3	4.362	-0.067	0.213	-0.007	-0.024	0.556	1.921

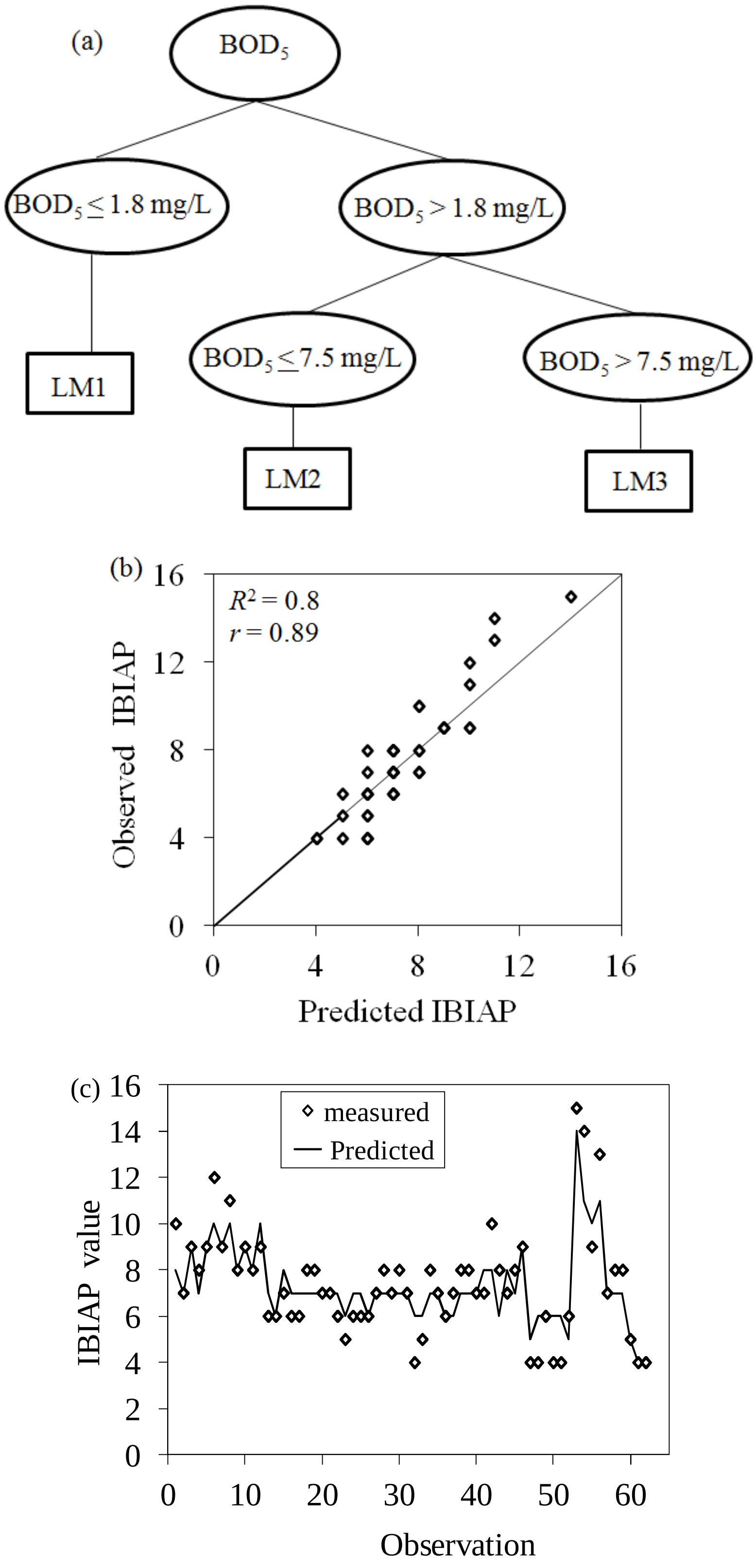


Fig. 4.5. Model tree (MT) relating the ecological water quality, assessed through the biotic (IBIAP) index, with basic physicochemical and hydraulic variables in the River Cuenca and its tributaries (a) MT and knowledge rules, (b) scatter plot, (c) analysis of goodness of fit.

4.3.4 Integrated ecological modelling and scenarios assessment

Using the developed integrated ecological model, the impact of the wastewater management plans on the physicochemical and ecological water quality of the Cuenca river and its tributaries was evaluated. Profiles of average concentrations of DO, BOD_5 and Faecal Coliforms (FC), were made for each pollution control scenario and each river considering the results obtained with the QUAL2Kw model. An example of the physicochemical predictions for each scenario in the main stem (Tomebamba and Cuenca rivers) is presented in Fig. 4.6a to 4.6c. Predictions of the presence/absence of the two target macroinvertebrate taxa and IBIAP values and classes for each monitoring station are presented in Table 4.4 and Fig. 4.6d. The impact of the sanitation plans on the ecological water quality is most clear in the Tarqui (Ta4 and Ta5 in Table 4.4) and Yanunkay rivers (Y3 and Y4 in Table 4.4) and along the main stem, especially in the River Cuenca (station C1 in the abscissa 17.5 km and C6 in the abscissa 27.5 km). Most of the wastewater management plans only took the urban area close to these rivers into consideration.

The results of the simulation of scenarios 1 and 2 show the importance of having a treatment for the domestic wastewater (i.e. WWTP) generated by the city of Cuenca. The added value of the WWTP becomes clear in the second scenario. Without treatment of the wastewater the BOD_5 and FC reach maximum values of 7.5 mg O_2/L and 2.3×10^5 MPN/100 mL, respectively at the last monitoring station of the River Cuenca (Fig. 4.6). The ecological water quality (EWQ) in scenario 1 shows lower IBIAP values compared with scenario 2 (Table 4.4 and Fig. 4.6). In Scenario 1, four of the six sampling sites of the River Cuenca (i.e. C2-C5) have IBIAP values of seven (moderate EWQ) and the last sampling site (i.e. C6) has a value of six (poor EWQ).

Scenario 2 (Fig. 4.6), shows the improvement in the water quality of the River Cuenca by treating the wastewater. This is reflected by BOD_5 concentrations between 3 and 4 mg O_2/L and FC between 6×10^4 and 8×10^4 MPN/100 mL (between stations C1 and C6). According to Chapman (1996), unpolluted waters typically have BOD_5 values of 2 mg O_2/L or less and FC of 1×10^2 MPN/100 mL or less, whereas those receiving wastewaters may have BOD_5 values up to 10 mg O_2/L or more and FC up to 1×10^7 MPN/100 mL or more, particularly near to the point of the wastewater discharge. DO is not a problem in this river, because the minimum value is always above 80% of DO saturation, which is about 6.7 mg/L. This is because of the high reaeration rates which are typical in mountain rivers with high turbulence and high flow velocities. Regarding the EWQ of the River Cuenca in scenario 2 (Table 4.4 and Fig. 4.6), the IBIAP has a value of eight in five of the six sampling sites (i.e. C1-C5) and a value of seven in the last station (C6), resulting in a categorisation of moderate EWQ in both cases. The absence of pollution sensitive taxa such as Trichoptera and the presence of pollution tolerant taxa such as Physidae in all the sampling sites of this river, shows the negative impact of the discharge of untreated wastewater.

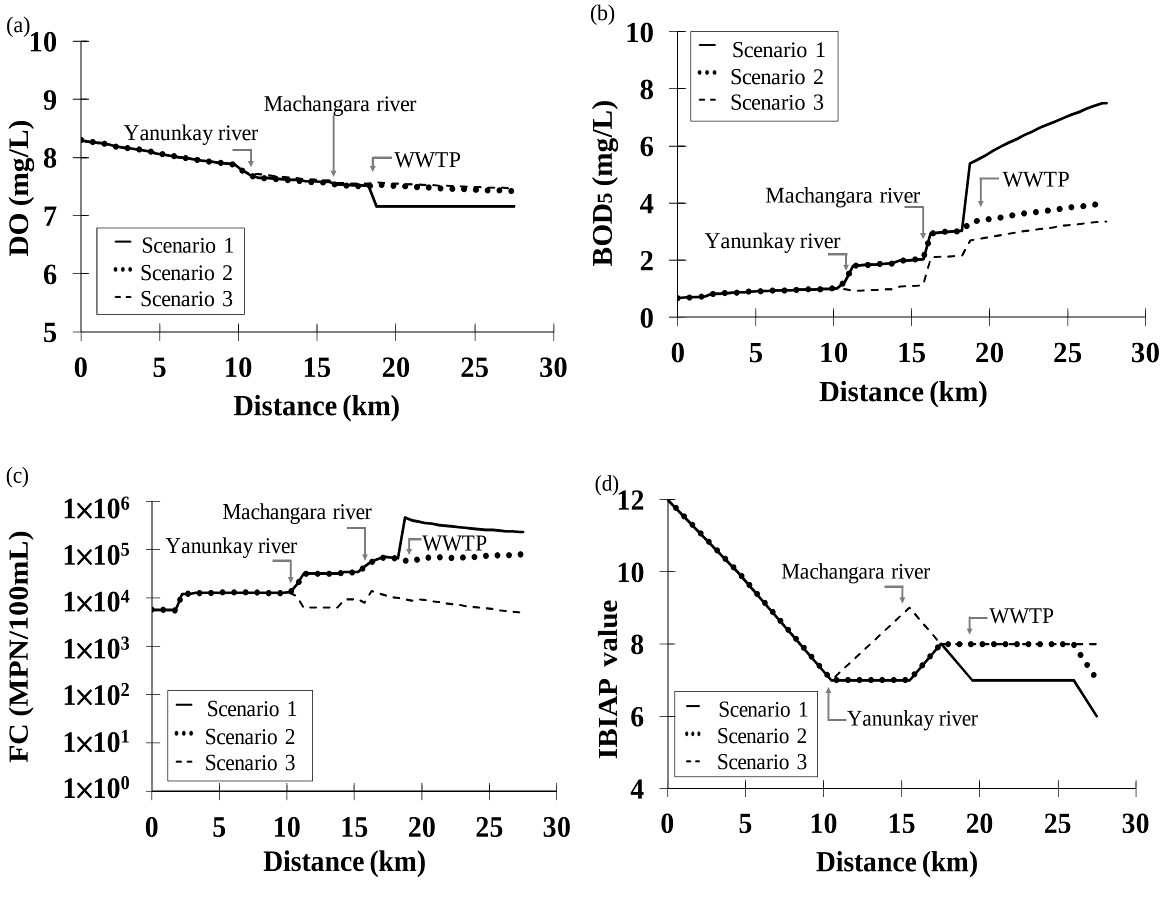


Fig. 4.6. Predictions of the dissolved oxygen (a), five-day biochemical oxygen demand (b), Faecal Coliforms (c) and IBIAP values (d) along the joint of the Tomebamba (between abscissa 0.0 km and 17.5 km) and Cuenca river (after abscissa 17.5 km) for the three different pollution control scenarios (Scenario 1: discharge of untreated wastewater; Scenario 2: functioning of the wastewater treatment plant; Scenario 3: collection and treatment of the total amount of domestic wastewater. Distance measured from the first station).

Table 4.4. Impact of different pollution control scenarios on the ecological water quality of the river Cuenca and its tributary rivers, expressed as the biotic (IBIAP) index and the presence/absence of macroinvertebrates. (Trich.: Trichoptera; Phys.: Physidae; Ta: Tarqui river; Y: Yanunkay river; Ma: Machangara river; Tb: Tomebamba river; C: Cuenca river; Scenario 1: discharge of untreated wastewater; Scenario 2: functioning of the wastewater treatment plant; Scenario 3: collection and treatment of the total amount of domestic wastewater).

Site	Scenario 1				Scenario 2				Scenario 3			
	Trich.	Phys.	IBIAP		Trich.	Phys.	IBIAP		Trich.	Phys.	IBIAP	
			Value	Class			Value	Class			Value	Class
Ta3	1	0	9	Moderate	1	0	9	Moderate	1	0	9	Moderate
Ta4	1	1	8	Moderate	1	1	8	Moderate	1	0 *	9 *	Moderate
Ta5	0	1	7	Moderate	0	1	7	Moderate	1 *	0 *	9 *	Moderate
Y1	1	0	12	Moderate	1	0	12	Moderate	1	0	12	Moderate
Y2	1	0	12	Moderate	1	0	12	Moderate	1	0	12	Moderate
Y3	1	1	10	Moderate	1	1	10	Moderate	1	0 *	12 *	Moderate
Y4	0	1	8	Moderate	0	1	8	Moderate	1 *	0 *	12 *	Moderate
Ma1	1	0	10	Moderate	1	0	10	Moderate	1	0	10	Moderate
Ma2	1	0	9	Moderate	1	0	9	Moderate	1	0	9	Moderate
Ma3	0	1	6	Poor	0	1	6	Poor	0	1	6	Poor
Ma4	0	1	6	Poor	0	1	6	Poor	0	1	6	Poor
Tb1	1	0	12	Moderate	1	0	12	Moderate	1	0	12	Moderate
Tb2	1	0	10	Moderate	1	0	10	Moderate	1	0	10	Moderate
Tb4	1	0	7	Moderate	1	0	7	Moderate	1	0	7	Moderate
Tb6	1	1	7	Moderate	1	1	7	Moderate	1	0 *	9 *	Moderate
C1	0	1	8	Moderate	0	1	8	Moderate	0	0 *	8	Moderate
C2	0	1	7 **	Moderate	0	1	8	Moderate	0	0 *	8	Moderate
C3	0	1	7 **	Moderate	0	1	8	Moderate	0	0 *	8	Moderate
C4	0	1	7 **	Moderate	0	1	8	Moderate	0	0 *	8	Moderate
C5	0	1	7 **	Moderate	0	1	8	Moderate	0	0 *	8	Moderate
C6	0	1	6 **	Poor	0	1	7	Moderate	0	0 *	8 *	Moderate

Notation for logistic regression model for Trichoptera and Physidae taxa: 1. Present, 0. Absent

The range of the IBIAP lies between 1 and 16, and the higher the value, the better the ecological water quality

* Ecological water quality improvement considering the reference situation (Scenario 2)

** Ecological water quality deterioration considering the reference situation (Scenario 2)

The results of the third scenario showed that the collection and treatment of the total amount of domestic wastewater generated by the city of Cuenca, allowed reaching values of 5.0×10^3 MPN/100 mL of FC and 3.4 mg O_2/L of BOD_5 at the end of the Cuenca river (Fig. 4.6). However, even in this scenario the maximum threshold values of 6×10^2 MPN/100 mL of FC for using this water for human consumption after conventional treatment (Ecuadorian legislation, TULAS (2002)), is not reached neither in the

Tomebamba nor in the Cuenca river. Regarding the concentrations of BOD_5, only the River Cuenca exceeds the maximum threshold value of 2 mg O_2/L of BOD_5 for this water use. Concerning the maximum threshold value of FC to guarantee the preservation of flora and fauna in the Cuenca and Tomebamba rivers, it can be seen that both rivers exceed the maximum concentration of 2×10^2 MPN/100 mL of FC (TULAS, 2002). Regarding the EWQ, there is an increase of the IBIAP values for this scenario compared with the reference situation (Scenario 2) in the stations Ta4 and Ta5 of the Tarqui river, Y3 and Y4 of the Yanunkay river, Tb6 of the Tomebamba river and C6 of the River Cuenca. Additionally, the presence of Trichoptera in Ta5 and Y4 and absence of Physidae in Ta4, Ta5, Y3, Y4, Tb6 and from C1 until C6, show the improvement in the EWQ.

4.4 Discussion

4.4.1 Integrated ecological modelling approach

Considering the limitations of the HSI approach and the WFD-Explorer, the IEMF was implemented and evaluated. This framework, integrates a detailed physical habitat and water quality model with data-driven models developed to predict the specific habitat conditions of aquatic species. The integrated model allowed assessing simultaneously the impact of physicochemical pollution and hydromorphological disturbances on the prevalence of macroinvertebrates and the ecological water quality. Habitat preference of macroinvertebrates is undoubtedly determined by multivariate processes where the preference for a location is based on several interacting variables (De Pauw and Hawkes, 1993; Goethals, 2005). Therefore, the use of multivariate approaches, such as the ones developed here, are more appropriate for the analysis of species-environment associations (Ahmadi-Nedushan et al., 2006). These types of approaches inherently consider the interrelation and correlation structure of the environmental variables (Ahmadi-Nedushan et al., 2006). Additionally, they allow identifying the physicochemical or hydromorphological characteristics that are the overriding factors to define the ecological water quality of rivers.

For the present study, the modular approach for model integration was selected. This approach integrates the model QUAL2Kw (Pelletier et al., 2006) with two ecological models (LRM and MT). These regression models were used to determine the relationship between a system's inputs and outputs using a training dataset that is representative for the energy fluxes within the ecosystem. Thus, LRM and MT allowed having ecological models in which direct relations between a set of predictor variables is calculated, without incorporating feedback loops. Daily average data generated by the QUAL2Kw model were used for the model integration during the scenario analysis. The validation of this approach was discussed in section 1.1 in Chapter 1.

4.4.2 Ecological water quality modelling of the River Cuenca

In an applied sense, (ecological) models are most useful as prediction tools and not only for exploring relationships in a historical dataset (Rushton et al., 2004). The proposed integrated ecological model allows modelling and assessing the ecological impact of wastewater discharges in the River Cuenca. Additionally, it can help to calculate the needed reductions in wastewater discharges of organic matter to meet biological quality criteria in this river. This integrated model can show the presumed impact of collecting all wastewater generated in the city of Cuenca and improving the wastewater treatment system. This approach should allow preserving habitats and species, to stop degradation and to improve river water quality. Models able to predict the habitat requirements of organisms help to ensure that planned actions meet the required effects for the target ecosystems.

The results of the models showed that the most important environmental variables to assess and predict the ecological water quality (EWQ) were BOD_5 (LRM model for Trichoptera and MT for IBIAP) and Faecal Coliforms (LRM model for Physidae). These results suggest that physicochemical indicators of organic pollution, such as BOD_5 and FC, are the overriding factors to define the EWQ in the rivers situated in the city of Cuenca.

4.4.3 Model performance

The performance of the QUAL2Kw was assessed by the R^2 and the *dm*. During the calibration process the *dm* and R^2 values were above 0.81 and 0.87, whereas during the validation the values were above 0.71 and 0.72, respectively. Those values represent that the model predicts the calibration and validation data relatively well. According to Harmel and Smith (2007) the *dm* is better suited to evaluate model goodness-of-fit than the R^2, because R^2 is insensitive to additive and proportional differences between model simulations and observations. However, one drawback of the *dm* is that, in general, it is more difficult to achieve high values, which makes it less attractive as efficiency criterion at first view (Krause et al., 2005).

For the validation of the ecological models a cross validation technique was used. This technique is particularly useful when only a limited number of data are available for training and validating the model (Gabriels et al., 2007). Unfortunately, partitioning the existing data is not a perfect solution since it is less efficient than collecting new data. In addition, the inevitable reduction in the size of a training set will usually produce a corresponding decrease in the sub-model accuracy. Therefore, a trade-off exists between having a large test set that gives a good assessment of the sub-model performance and a small training set that is likely to result in a lower performance (Fielding, 2002). This analysis can be seen in Table 4.2 in which the performance indicators for the LRM built for Physidae and Trichoptera with the complete dataset were relatively higher than those obtained with only the training set or the test set when the three-fold cross validation process was applied.

Current practice in species distribution modelling suggests applying at least two different performance criteria for model evaluation (Mouton et al., 2010). Threshold-dependent approaches such as CCI and *K* have received some criticism because they are affected by prevalence (Fielding and Bell, 1997; Fielding 2002; Rushton et al., 2004; Tirelli et al., 2009; Mouton et al., 2010). Thus, the use of threshold-independent approaches such as the area under the curve for a ROC plot (AUC), has been increasingly used in the assessment of logistic regression models (Pearce and Ferrier, 2000; Manel, et al., 2001; Guisan, 2002; Fielding, 2002; Willems et al., 2008). However, some authors suggested that the AUC appears to be independent of prevalence only in its middle range (Maggini et al., 2006; McPherson and Jetz, 2007). Maggini et al. (2006) found that the AUC is systematically lower at extreme prevalence values (prevalence <0.05 or >0.70). Considering these advantages and drawbacks of the three performance indicators (CCI, *K* and AUC), it was decided to use all of them in order to select the best model.

The models presented in this study can still be improved in some aspects. The LRM and MT were only based on data from 60 samples. To optimize the models, more data should be collected in surface waters characterized by a high or good ecological water quality and more variables need to be monitored in a consistent way. Therefore, the water quality model QUAL2Kw should predict some additional variables so that these can be included in the regression models, such as conductivity, particulate inorganic and organic matter (e.g. Inorganic Suspended Solids and detritus) and nutrients (i.e. different status of Nitrogen and Phosphorous). A better coordination of the monitoring networks and encodings can yield a more comprehensive dataset and more reliable and ecological relevant models. The data collection strategy should focus on datasets where all variables are gathered during each sampling event, especially with regard to the flow variables. Besides this, the model can also be improved technically. For instance, the reliability of the LRM could be improved by the application of prevalence adjusted optimisation and the combination of data-driven and knowledge based models (Mouton et al., 2009b; 2009c).

4.4.4 Using integrated modelling for decision support in water quality management

The results proved that integrated models like the one presented here give an added value for decision support in water quality management by coupling ecological quality to a set of hydraulic and physicochemical water quality measures based on the simulation model QUAL2Kw. The application of the integrated ecological modelling showed that the LRM and MT helped to consider receiving water's ecological aspects in the wastewater treatment/disposal strategies of the different scenarios. Any improvement in the EWQ in a monitoring station, was represented by an increase of the biotic index (IBIAP) and the presence of pollution sensitive taxa (i.e. Trichoptera) or absence of pollution tolerant (i.e. Physidae) taxa.

The simulation of scenarios for wastewater management in the city of Cuenca, suggest that the collection and treatment of all the domestic wastewater generated by the city (scenario 3), is not enough to achieve a good EWQ in the River Cuenca and its tributaries. It was considered that it is necessary to control the impact of the industrial wastewater discharges, because up to date there is no sanitation plan for reducing this impact. The organic pollution of the industrial wastewaters is similar to the domestic wastewaters which are not yet treated (1.1 tons of organic matter per day in terms of BOD_5). Additionally, diffuse pollution, such as wastewater discharges from agricultural activities and scattered houses, should be controlled before the first monitoring station of the Tomebamba river (i.e. Tb1) and at the upper catchment of the Yanuncay and Tarqui rivers. The pollution control in this area could allow reaching BOD_5 concentrations and FC values that indicate a low human impact and that allow improving the EWQ.

4.5 Conclusions

In this study, the proposed IEMF was tested on a case study in the River Cuenca in Ecuador, with the integration of the hydraulic and physicochemical water quality model QUAL2Kw and two ecological models. These ecological models allow predicting the presence of two target taxa of macroinvertebrates, Trichoptera (pollution sensitive taxon) and Physidae (pollution tolerant taxon) and the value of the biotic index IBIAP. The integrated ecological model was used to simulate three scenarios for water management plans in this river. The two ecological models clearly indicated an increase in potential habitat availability for Trichoptera and a decrease in this potential habitat for Physidae, as the pollution load from domestic and industrial wastewaters is reduced.

Chapter 5: Case study 3: Assessing the ecological impact of upgrading an existing wastewater treatment plant on the Drava River in Croatia

Adapted from:

Holguin-Gonzalez J.E., Boets P., Everaert G., Pauwels I., Lock K., Gobeyn S., Benedetti L., Amerlinck Y., Nopens I., Goethals P.L.M. (submitted). Development and assessment of an integrated ecological modelling framework for decision support in water management. Water Resources Management.

Holguin-Gonzalez, J.E., Everaert, G., Benedetti, L., Amerlinck, Y., Goethals P.L.M. (2012). Use of habitat suitability modeling in the integrated urban water system modeling of the Drava river (Varaždin, Croatia). In: H. Mader & J. Kraml (Eds.). Proceedings of the 9th International Symposium on Ecohydraulics-ISE 2012. Vienna, Austria, September 2012. ISBN: 978-3-200-02862-3.

Holguin-Gonzalez, J.E., Everaert, G., Benedetti, L., Amerlinck, Y., Goethals P.L.M. (2012). Integrated ecological modelling for decision support in the integrated urban water system modelling of the Drava river (Varaždin, Croatia). In: R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.). International Environmental Modelling and Software Society (iEMSs), Sixth Biennial Meeting. Proceedings of the International Congress on Environmental Modelling and Software, "Managing Resources of a Limited Planet: Pathways and Visions under Uncertainty", Leipzig, Germany, July 2012. ISBN: 978-88-9035-742-8.

Boets, P., Holguin-Gonzalez, J.E., Lock, K., Goethals, P.L.M. (2013). Data-driven habitat analysis of the Ponto-Caspian amphipod *Dikerogammarus villosus* in two invaded regions in Europe. Ecological Informatics 17, 36-45. doi:10.1016/j.ecoinf.2012.07.001

Chapter 5: Case study 3: Assessing the ecological impact of upgrading an existing wastewater treatment plant on the Drava River in Croatia

Abstract:

The aim of this study was to evaluate a conceptual framework for model integration (IEMF, presented in Chapter 1) towards decision support in an integrated urban drainage system located in the city of Varaždin, Croatia. Based on the integrated modelling framework, the effect of upgrading a wastewater treatment plant (WWTP) on the ecological state of the receiving river in an urban drainage system was assessed. The IEMF integrated four models, being a model assessing the WWTP processes, a model simulating the river water quantity, a model predicting the physicochemical water quality and finally a model assessing assess the ecological river water quality. Three potential investment scenarios of the wastewater treatment infrastructure in the city of Varaždin (Croatia) were implemented and their impact on the ecological water quality was assessed. From this scenario-based analysis it was concluded that upgrading the existing WWTP, with nitrogen and phosphorous removal, will not be sufficient to reach a good ecological water quality in the Drava river which is receiving the effluent of the WWTP. Therefore, addition point and diffuse pollution sources in the area should be monitored and remediated. The ecological models developed helped identifying that the impact of physicochemical pollution on the river ecology, generated by the discharge of wastewaters, is significantly influenced by local conditions of water velocity, water depth, type of substratum and channel morphology (i.e. hydromorphological conditions).

5.1 Introduction

Integrated water management requires an understanding of the elements that affect the ecological state of a river system and enables to predict how these will respond based on different management options. Most traditional modelling frameworks are not able to meet these requirements as models tend to represent individual processes and to run independently (Kraft, 2011). Thus, integrated modelling frameworks are required. These integrated frameworks allow performing comprehensive evaluations which would not be possible when investigating each individual component of the system separately. The integration of models is the key for success as integrated models can be more efficiently applied in environmental decision making.

Traditionally, investments in sanitation infrastructure of urban wastewater systems have been assessed considering the fulfilling of legal physicochemical emission limits without considering the ecological state of the receiving waters (Devesa et al., 2009). Countries which are in the process to join the European Union (EU) should fit their legislation to EU standards, including the European Water Framework Directive (WFD). The WFD promotes the integrated approach in river management, considering the concept of ecological state. This state refers to the quality of the structure and functioning of the aquatic ecosystem of the surface water. It is defined in terms of the quality of the biological community and the hydromorphological and physicochemical characteristics. Furthermore, the WFD promotes a combined approach of the emission limit values and the recipient quality standards and encourages the availability and use of decision support tools for water management (Devesa et al., 2009). Future investments in the construction of new municipal wastewater treatment plants (WWTPs) and in the upgrading of existing WWTPs (secondary and tertiary treatment) are planned in the coming years in several European countries. Therefore, the development and application of integrated ecological modelling tools to assess the impact of these investments on the ecological state of the receiving waters are necessary.

Two of the most important pressures that determine the ecological river water quality are hydromorphological disturbances and physicochemical pollution. The integration of mathematical models in water management allows analyzing these two types of pressures. Current practice in model integration focuses on hydromorphological pressures using hydrological or hydraulic modelling and habitat suitability methods for two main purposes: (1) to identify flow regimes for ecological protection (e.g. USGS, 2001; Hughes and Louw, 2010; Paredes-Arquiola et al., 2011; Jähnig et al., 2012); (2) to design and to evaluate river restoration schemes (e.g. Bockelmann et al., 2004; Tomsic et al., 2007; Everaert et al., 2013). The most widely known and applied hydraulic habitat simulation software is the Physical Habitat Simulation Model, PHABSIM (Bovee et al., 1998; USGS, 2001), a component of the Instream Flow Incremental Methodology (IFIM) (Stalnaker et al., 1995). IFIM was developed to integrate aspects of instream flow problems, including the water needs of aquatic ecosystems (Stalnaker et al., 1995). PHABSIM predicts how the physical habitat (e.g. depth, velocity, substrate) depends on flow regime and combines this information with habitat suitability criteria to determine a suitability index for a given species as a function of flow (e.g. fish and macroinvertebrates) (Bovee et al., 1998). Unfortunately, PHABSIM does not directly address other elements of stream ecosystems such as water quality and energy inputs (USGS, 2001). However, when the impact of physicochemical pollution, such as wastewater discharges, is the main factor determining the ecological river water quality, the application of methods based on hydraulic habitat simulations only cannot properly assess the effects. The WFD-Explorer toolbox (Deltares, 2009) is good attempt to deal with this interaction of environmental variables, however, as it was mentioned before (in Chapters 3 and 4) this software has some limitations for its use in a small scale and its application outside The Netherlands (Mouton et al., 2009a).

Considering the limitations and simplifications of software packages such PHABSIM and the WFD-Explorer, the IEMF presented in Chapter 1 was implemented. In this study, all four basic modelling components of the IEMF were considered (see Fig. 1.1, in Chapter 1). The first model corresponds to the simulation of wastewater treatment plant (WWTP) processes, the second deals with river water quantity, the third considers physicochemical river water quality and the fourth corresponds to a river ecological assessment model. The proposed framework includes a detailed physical habitat and water quality model linked to ecological models based on abiotic river conditions. This integrated approach allows assessing simultaneously the impact of hydromorphological pressures and physicochemical pollution (e.g. discharge of a WWTP) on the ecological river water quality (as requested by the WFD). This study describes the implementation and evaluation of the IEMF in a Croatian river (Drava river) as a tool for decision support in an integrated urban drainage system located in the city of Varaždin, Croatia. Three scenarios evaluating the effect of upgrading the existing WWTP, with nitrogen and phosphorous removal, on the ecological water quality (EWQ) of the receiving river were assessed. The model simulating the processes of the WWTP (Activated Sludge Model No. 2d (ASM2d); Henze et al., 2000) was implemented in the simulation platform WEST (World wide Engine for Simulation, Training and Automation, Vanhooren et al, 2003). The Drava river water quantity and quality were modelled with the River Water Quality Model No.1 (RWQM1, Reichert et al., 2001a) developed with a Matlab (Matrix Laboratory 7.10; MathWorks, 2010) application. For the ecological modelling an ecological assessment model for rivers based on regression trees (Breiman et al., 1984) was built in Matlab. This data-driven modelling approach allowed predicting the EWQ of the Drava river.

5.2 Materials and methods

5.2.1 Study area

The Drava river is a transboundary river that springs in Italy at an altitude of 1192 meters above sea level and runs, for almost 730 km, through five countries (Italy, Austria, Slovenia, Croatia and Hungary). This study focuses on the Drava river stretch located in the north-east part of Croatia, in the Varaždin County (Fig. 5.1). In this zone, the Drava river consists of a succession of three lakes (i.e. reservoirs) called lake Varaždin, lake Čakovec and lake Dubrava. From every lake, part of the Drava river is diverted to a hydroelectric power plant (HPP) through a tailrace canal, while the remaining water is released through a dam in the old Drava's river path. This is an example of a multifunctional river ecosystem which has been heavily modified in order to exploit resources and services, mainly hydroelectricity generation. Remnants of the original meandering river channel between the dams still remain and support a rich nature, however, the main stream flow goes through the hydroelectricity generation system. The fragmentation resulting from the existing dams and the unnatural daily flood wave from the electricity generating cycle has major impacts on the migration of fish and their spawning, resulting in a large decline of the fish population (WWF, 2003). Additionally, the stream flow reduction affects the dilution or self-cleaning capacity of the river, especially near the city of Varaždin (after lake Čakovec), where treated and untreated wastewaters from agricultural, urban and industrial activities are discharged into the river. Industrial activities such as milk and meat production have been identified as the main driver of pollution pressures in the system (VARKOM et al., 2010). These activities need energy in order to manufacture goods and to provide services. The need for energy drives the competition between the quantity of water available for electricity production and the amount of water released to the river system available for natural dilution.

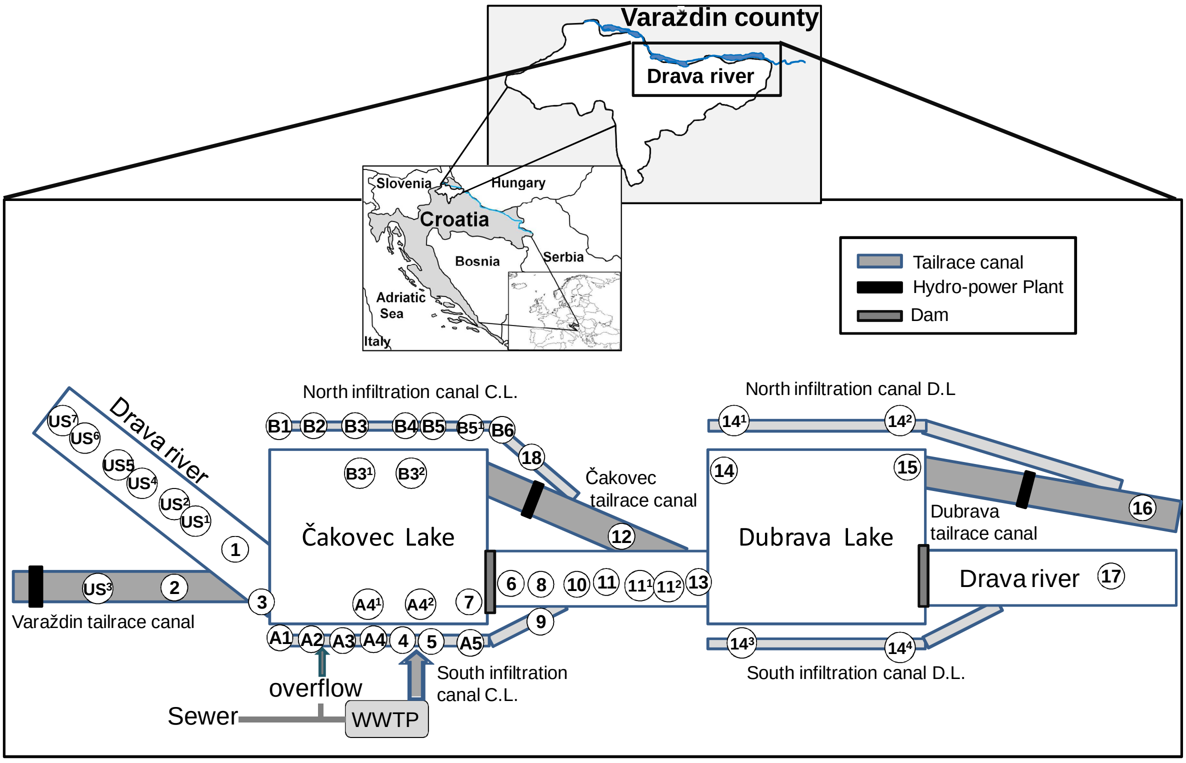


Fig. 5.1. Overview of the study area and scheme of the studied system with indication of the Drava river and sampling sites in the Varaždin county, Croatia. (C.L.: Čakovec lake; D.L.: Dubrava lake).

Besides industrial activities, the discharge of domestic wastewater from the city of Varaždin, with around 50,000 inhabitants, is the second main source of pollution in the study area (VARKOM et al., 2010). The water distribution and sewage system management company in Varaždin (VARKOM) has been operating since the year 1962 to improve the water quality of the Drava river. This company has been investing in infrastructures used for environmental protection, such as the collection and treatment of wastewater. Varaždin has a combined sewerage flowing into a WWTP that treats both municipal and industrial (mainly food processing industry) wastewater. The plant was originally designed for carbon removal only. In spite of the daily-averaged good effluent quality, the WWTP has some problems such as the lack of treatment for the overflow of the sewer system (i.e. in extreme rain events) and the overload of the secondary clarifiers (resulting in high suspended solids concentrations in the effluent) (VARKOM et al., 2010). The effluent of the WWTP ends up in a small canal (i.e. south infiltration canal C.L.) that collects the infiltration water from lake Čakovec and some streams (that joint the small canal) and finally it joints the Drava river after the lake (station 9 in Fig. 5.1).

5.2.2 Data collection, coupling of data and dataset pre-processing

The dataset used in this research results from the information collected during three monitoring campaigns made in the framework of the project Water Treatment Optimization with Ecological Criteria-WATROPEC (VARKOM et al., 2010) (i.e. April and October 2010) and by the authors (i.e. September 2011). The monitoring campaigns allowed collecting simultaneous information about physicochemical and hydromorphological conditions, physical habitat conditions and macroinvertebrate composition of 103 records collected from 60 sampling locations. Additionally, information related to municipal wastewater production and daily and annual average data for the WWTP influent, effluent and sludge were collected.

The physicochemical assessment in the Drava river included three variables measured in the field being dissolved oxygen (DO, mg O_2/L), temperature (T, °C) and pH (-). The five-day biological oxygen demand (BOD_5, mg O_2/L), total nitrogen (TN, mg N/L), nitrate (NO_3, mg NO_3-N/L), ammonium (NH_4^+, mg NH_4^+-N/L), total phosphorus (TP, mg P/L), phosphate (PO_4, mg P/L), total suspended solids (TSS, mg/L), chemical oxygen demand (COD, mg O_2/L) were measured in the laboratory. Additionally, organic nitrogen (ORGN, mg N/L) and organic phosphorus (ORGP, mg P/L) were calculated based on the previous variables. For the hydromorphological assessment, average water depth, average water velocity and a categorical variable called 'Type' that holds information on the hydromorphological structure of the water body were considered. Two categories or levels were defined for this variable: (1) hydromorphological favourable (value of one): natural bank structure, mixed bottom substrate, thin sludge layer, meandering, heterogeneous bank and bottom structure; and (2) hydromorphological unfavourable (value of two): artificial bank structure, tick sludge layer, straight waterway, homogeneous bank and bottom structure. For the biological assessment, macroinvertebrates were sampled by hand net as described by Gabriels et al. (2010). Identification was carried out according to the taxonomic levels defined by Gabriels et al. (2010) needed to calculate the biotic index, this means family or genus level. The river state in each sampling location was estimated according to the ecological quality ratio (EQR), ranging from 0 to 1, using the Multimetric Macroinvertebrate Index Flanders (MMIF; Gabriels et al., 2010). This index is calculated based on the occurrence and abundance of macroinvertebrate taxa and their sensitivity to organic pollution. The MMIF is a type-specific multimetric index based on five equally weighted metrics: taxa richness, number of Ephemeroptera, Plecoptera and Trichoptera (EPT) taxa, number of other sensitive taxa, the Shannon-Wiener diversity index and the mean tolerance score. In the context of the WFD and for transparency towards decision makers, the EQRs are converted to five ecological quality classes: bad (0-0.3), poor (0.3-0.5), moderate (0.5-0.7), good (0.7-0.9) and excellent (0.9-1.0). The MMIF, developed for Flanders (northern part of Belgium), was used in this study since this method is generally applicable and can be used in other countries and rivers with similar characteristics (Lock et al., 2011).

In order to couple the ecological model with the Matlab application and WEST, a dataset of 103 records containing simultaneous measurements (based on sampling location and time) of physicochemical, hydromorphological and biological variables was compiled. These variables were selected considering that the IEMF will be used to evaluate the effect of upgrading the WWTP to tertiary treatment (which implies carbon and nutrient removal). In total, ten predictor variables (DO, BOD_5, NO_3, NH_4^+, ORGN, PO_4, ORGP, average water depth, average water velocity and the hydromorphological 'Type') and one response variable (MMIF index as a continuous value) were selected (see Appendix B; Table B.3).

The data available for building the ecological models was pre-processed following the procedure described in sections 3.2.2 and 4.2.2. Firstly, an evaluation of outliers was performed using two graphical tools, box plots and Cleveland dot plots (Zuur et al., 2010). Additionally, a preliminary mass balance analysis to evaluate the reliability of the data and possible outliers was performed with the water quality model. This analysis showed that there were seven possible outliers. However, both analysis with and without outliers were considered during the building procedure for the ecological models, to assess the impact of the outliers in the ecological models building procedure. Details about the dataset pre-processing of the Drava river are presented in Appendix E1. Secondly, collinearity between the predictor variables was assessed by a Principal Component Analysis (PCA) with a varimax rotation, to maximise the independence of the Principal Components and the Spearman rank (*S*) correlation coefficient. The correlation matrix and PCA helped determining the correlation between the potential predictor variables. Thirdly, relationships between the response variable (MMIF index) and the predictor variables were evaluated with the (*S*) correlation coefficient.

Regarding the WWTP, the variables measured in the influent and effluent were: flow rate, temperature (air and water), pH, DO, TSS, COD, BOD_5, NH_4^+, NO_3, TP, PO_4 and chlorides, whereas in the sludge the following variables were measured: TSS, sludge volume index, organic and mineral content.

5.2.3 Model building, validation and implementation

5.2.3.1 Wastewater treatment plant model

A model characterising the processes in the WWTP was implemented in WEST, which is a modelling and simulation software platform for biological wastewater treatment systems that incorporates processes such as carbon oxidation in aerobic and anaerobic conditions, nitrification, denitrification and phosphorus removal (Vanhooren et al., 2003). The WWTP processes were modelled using an adaptation of the Activated Sludge Model No. 2d (ASM2d; Henze et al., 2000), to allow different decay rates under different environmental conditions (Gernaey and Jørgensen, 2004). The WWTP designed for carbon removal only, treats a combined sewerage flow with both municipal and industrial (mainly food processing industry) wastewater. The initial part of the system comprises of a screening system and a long (aerated) channel, with three overflows allowing a maximum influent to the biological stage of 500 m^3/h. The biological stage comprises of two parallel lanes. The South lane exists of an Integrated Fixed-film Activated Sludge (IFAS) System followed by a conventional activated sludge tank. The North lane exists of two conventional activated sludge tanks. The water leaves the plant via rectangular secondary clarifiers (VARKOM et al., 2010). The WWTP model was calibrated and validated with daily and annual average data for the WWTP influent, effluent, sludge and information related to municipal wastewater production. Details about the implementation of the model characterising the processes in the WWTP are described by VARKOM et al. (2010).

5.2.3.2 Hydraulic and physicochemical river water quality model

The hydraulic and physicochemical river water quality models were developed based on a certain river stretch, located near the city of Varaždin, where the main impacts of physicochemical pollution and hydromorphological pressures on the ecological river quality are identified. This modelling stretch included: (1) lake Čakovec (C.L.); (2) the south infiltration canal C.L. with inputs of the combined sewer overflow, the WWTP and inputs of untreated wastewater; (3) the Drava river (succession river-lake-river) with inputs of the Varaždin tailrace canal and the south infiltration canal C.L.

To model the water flow of the system, two methods can be used: the complex hydraulic routing method solving the 'St.Venant' equations (De St. Venant, 1871) and the conceptual hydraulic routing method (Deksissa et al., 2004). If the river system is not affected by backwater and tidal effects, such complex hydraulic model can be simplified by using a surrogate such as a conceptual hydraulic model. Previous studies (Camacho and Lees, 1999; Deksissa et al., 2004; Deksissa, 2004; Benedetti et al., 2007) have investigated the use of hydraulic surrogate models. In this study, the hydraulics were modelled by following a Continuous Stirred Tank Reactor in Series (CSTRS) approach (Whitehead et al., 1979; Beck and Reda, 1994; Deksissa et al., 2004). The CSTRS approach combines the continuity equation with an analytical or empirical relationship between the storage of water in the system (or reservoir) and the outflow. This approach requires an initial subdivision of the river into different stretches. These stretches were assumed to have uniform hydraulic and morphologic features; the section shape and discharge rating curve were assumed to be similar over these stretches. The information concerning average flow and average water height during the days of the monitoring campaigns provided by the Croatian Electricity Company (Grian and Kerea, 2004) was used to estimate the average velocity and width on several locations. Three gauging stations are situated within the study river stretch. For these sites flow-rating curves (Q-h relationships) of good quality were available. In other locations, where measures of water velocity were available, the average flow and water height provided were used to estimate the transversal area and river width (e.g. stations at lake Čakovec). The estimated width was compared with the estimated width in the GIS platform ARKOD available for free consulting by the Croatian Agency for payments in agriculture, fisheries and rural development (MAFRD, 2009). It was assumed that conditions of uniform steady flow were valid and backwater and tidal effects were not considered.

Two methods can be used to model river water quality processes: the complex pollutant transport routing, also known as advection-dispersion model and the conceptual pollutant transport routing (Deksissa et al., 2004). In this study, the second method was implemented. This method was based on the concept of using a cascade of CSTRS to represent the transport of pollutants through the Drava river and the infiltration and tailrace canals. Previous studies (Deksissa et al., 2004; Deksissa, 2004; Benedetti et al., 2007) have demonstrated the great potential of using the cascade of CSTRS approach in river water quality modelling. The River Water Quality Model No.1 (RWQM1, Reichert et al., 2001a) was implemented in the pollutant transport sub module. In order to use this sub module, a mass balance for a given finite time period was set up for every physicochemical variable. Details about the implementation of the water quality model are summarized in Appendix E2.

The calibration and validation of the hydraulic and physicochemical river water quality models were performed independently using the information collected during the monitoring campaigns of September 2011 (calibration) and April and October 2010 (validation). The hydraulic model was calibrated by changing two parameters, i.e. the Manning roughness coefficient of the river bed (n) and the slope of the river (S_0). Based on the available information and n values reported in the literature (Chow, 1981), initial conditions were proposed for these two parameters. Both parameters were adjusted in function of the simulations and measurements of the flow and water height of the considered stretch. The evaluation of the goodness of fit during the calibration and validation processes was performed by taking the difference between the estimated and the modelled uniform steady-state flow and water height. The calibration of the water quality model was performed by a constraint-based random search method. For this analysis, 1000 combinations of the calibration parameters (i.e. model rate parameters), considering values from uniform distributions, were evaluated with simulations. The calibration ranges of model rate parameters required (Appendix E2), were obtained from the literature (Chapra, 1997; Kannel et al., 2007; Cho and Ha, 2010). The calibration parameters were considered equal for every stretch in function of the type of water body (river, infiltration canal and lake). The evaluation criterion considered during the calibration and validation was the determination coefficient (R^2), which evaluates the goodness of fit between the simulations and the measurements. The R^2 values were calculated for each of the thousand simulations performed during the calibration and for each modelled variable. The model was calibrated separately for: (1) the Drava river; (2) the south infiltration canal C.L and (3) lake Čakovec. Modelling efforts were focused on DO, BOD_5, ORGP, PO_4, ORGN, NH_4^+, NO_3, average water depth and average water velocity.

5.2.3.3 Ecological model

Decision tree models are one of the most popular machine learning techniques used for ecological modelling because they are simple, transparent, easy to use and to interpret (Debeljak and Džeroski, 2011; Everaert et al., 2011; Gal et al., 2013; Boets et al., in press a). Decision trees are mostly used for predictive modelling and for extracting new knowledge about the observed processes. These models are generated through an iterative splitting of data into subspaces of the whole attribute space, where the goal is to maximize the distance between groups at each split (Stravs et al., 2008). Decision tree models, allow representing a series of rules that led to a result in the leafs of the model that can be: (1) class values (classification trees); (2) constant numerical values (regression trees); (3) linear equations (model trees).

In a preliminary assessment (i.e. without taking into account independent or internal validation; see details further) the applicability of the three types of decision tree models was evaluated based on prediction capacity and statistical reliability. Regression trees (RTs) gave the best results. Therefore, we used RTs with the classifier algorithm M5 (Quinlan, 1992; Wang and Witten, 1997) implemented in the statistical toolbox of Matlab (MathWorks, 2010). This algorithm is based on the classification and regression tree functions of Breiman et al. (1984). RTs were grown with a recursive partitioning algorithm from a training set of records, which is known as 'Top-Down Induction of Decision Trees' (Quinlan, 1986). For each step, the most informative input variable is selected as the root of the sub-tree and the current training set is split into subsets according to the values of the selected input variable. Subsequently, the dataset is split up in two sub datasets. This procedure is continued until a stop criterion is reached. The RT implemented allows performing a river EWQ assessment by predicting the value of the ecological MMIF index (Gabriels et al., 2010). This model uses physicochemical and hydromorphological variables (i.e. abiotic river conditions) as predictor variables. In RTs the development and the structure of the model allow the user to understand how each input variable contributes to the structure of the tree and to identify associations and general trends in the data. By implementing independent physicochemical and hydromorphological input variables and following the hierarchical structure of the tree, these tests lead to the associated predicted MMIF value.

Two main approaches exist for evaluating the predictive power of an ecological model. The first approach (independent validation) is to use two independent datasets, one for calibrating (training dataset) and another for evaluating the model (evaluation dataset). The second approach (internal validation) is to use a single dataset to calibrate the model and then evaluate it by resampling methods (Verbyla and Litvaitis, 1989), such as cross validation, leave-one-out- cross validation, also known as Jack-knife, or bootstrapping techniques. The RTs were built based on both approaches (i.e. independent and internal validation). In the first approach, the monitoring campaign of September 2011 was used for training and the monitoring campaigns of April and October 2010 were used for validation. In this approach, two datasets were considered, the first one with outliers (103 records) and the second one without outliers (96 records). The results of both analyses were compared in terms of statistical reliability and prediction capacity using four model performance criteria: (1) Pearson (r) correlation coefficient; (2) R^2; (3) root mean square error (RMSE) and (4) correctly classified instances (CCI).

In the second approach, a resampling method based on a bootstrapping technique (Verbyla and Litvaitis, 1989) was implemented using the dataset without outliers. The bootstrapping technique is an approach in which a smaller subsample of the available data was used to train and develop a model. Therefore, a subsample of the dataset was used for the tree construction, while the remaining part of the dataset was used for the validation (cross-validation). The subsample dataset for each of these models was based on stratified runs. A stratified dataset is a smaller dataset generated from the total dataset, which has the same number of instances of each MMIF class represented in the set (Everaert et al., 2013). In this study the MMIF classes were unequally represented in the dataset. The sampling stations at the Drava river basin had mainly bad EWQ (22 instances), poor EWQ (40 instances) and moderate EWQ (26 instances) and only few stations had a good EWQ (8 instances). Therefore, it was necessary to implement a stratification procedure for the dataset together with the bootstrapping technique in order to guarantee the same number of instances of each MMIF class represented in the set and to avoid biased models (Everaert et al., 2013). A subgroup of data with only excellent and good EWQ was created. Thus, the rest of the MMIF classes (i.e. moderate, poor and bad) in the stratified dataset had the same number of instances of the group of excellent plus good EWQ (Everaert et al., 2013). An example of the procedure followed is presented in Appendix E3. Bootstrapping methods allowed approaching the bias of an estimation by performing multiple resampling (with replacement) within the training dataset, and then to remove it to obtain an unbiased estimate (Efron and Tibshirani, 1993). This approach was repeated 1000 times, thus 1000 models were built and the model performance criteria r, R^2, RMSE and CCI were evaluated for each model. In the bootstrapping technique these performance criteria were tabulated and evaluated. The best 10 samples, in function of the CCI, were retained.

Recently, Larocque et al. (2011) and Everaert et al. (2012) stated that apart from the statistical reliability also the applicability and the ecological relevance are important aspects for model selection. Therefore, all models developed in both approaches (i.e. independent and internal validation) were ultimately assessed in three steps: (1) statistical reliability (r, R^2, RMSE and CCI); (2) ecological insight by including stakeholder's opinion and their expert knowledge to determine ecological relevance of the model. In case that biological inconsistencies were found, these models were dismissed; (3) applicability and practical use for decision support in water management.

5.2.4. Simulations of river management options

Using the IEMF three different wastewater treatment scenarios considering the upgrading of the WWTP with nutrient removal were evaluated. These scenarios were: (1) current situation, (2) upgrading of WWTP with nitrogen (N) and phosphorous (P) removal and; (3) upstream treatment and upgrading of WWTP with N and P removal. The RTs developed can be used to make predictions about the dependent variable (i.e. MMIF index) based on other independent values (i.e. physicochemical and hydromorphological variables) than the values that were used to build the model. Therefore, the physicochemical and hydraulic simulations of each scenario were used as input variables for the RTs. Daily average predictions of these input variables at each sampling station were considered (see discussion about this approach in section 1.1 in Chapter 1).

5.3 Results

5.3.1 Data analysis and variable selection

Regarding the PCA, the first six principal components (PCs) explained 77 % of the variance in the data and the variables that were included in the same PC were for PC1: TP, PO_4 and ORGP, for PC2: COD and BOD_5 and PC3: TN and ORGN. The correlation analysis between predictor variables showed seven highly-positively correlated variables (i.e. $S > 0.7$): TN with ORGN ($S = 0.89$); TP with PO_4 ($S = 0.82$) and TP with ORGP ($S = 0.77$) and; COD with BOD_5 ($S = 0.78$). The rest of the variables were not correlated ($-0.2 \leq S \leq 0.2$) or slightly correlated ($-0.45 \leq S \leq -0.2$ and $0.2 \leq S \leq 0.45$). DO ($S = 0.36$), PO_4 ($S = -0.22$) and NH_4^+ ($S = -0.19$) showed the highest correlation with the MMIF index, whereas average water velocity, TN and average water depth showed the lowest correlation with this index (S = 0.019, 0.017 and 0.002 respectively). The PCA and correlation analysis showed that the variables with a high degree of collinearity were: (1) BOD_5 and COD; (2) TN and ORGN and; (3) TP, PO_4 and ORGP. For constructing the ecological models highly correlated predictor variables were discarded. Hence, nine predictor variables were retained for the regression trees: DO, BOD_5, ORGN, NH_4^+, NO_3, ORGP, average water depth, average water velocity and hydromorphological type.

5.3.2 Hydraulic and physicochemical water quality model

The results of the calibration and verification processes of the water quantity and quality models showed that they reliably predict the trend and maximum and minimum values of DO, BOD_5, ORGP, PO_4, ORGN, NH_4^+, NO_3, average water depth and average water velocity in the river. As an example, the results of the calibrated model for the monitoring campaign of September 2011 (SC3, Sample campaign Nr. 3) are presented in Fig. 5.2 and 5.3. Similar graphs were built for the validation process and hydraulic variables, and some examples are presented in Appendix E2.

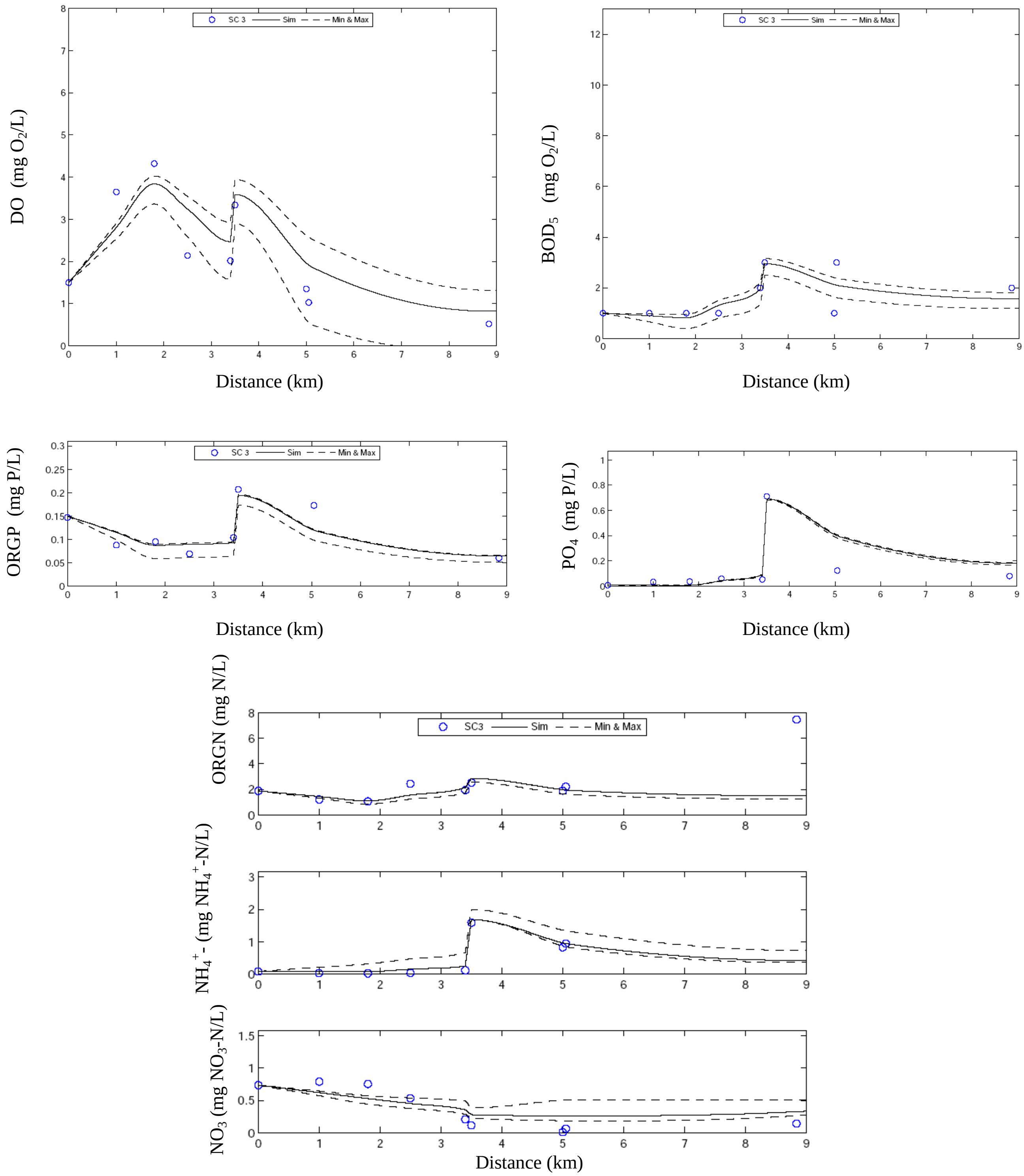


Fig. 5.2. Calibrated water quality model for dissolved oxygen (DO), five-day biological oxygen demand (BOD_5), organic phosphorus (ORGP), phosphate (PO_4), organic nitrogen (ORGN), ammonia (NH_4^+) and nitrate (NO_3) in the south infiltration canal of the Čakovec lake. The actual simulation is given by the continuous line. The dotted lines indicate the maximal and minimal simulated values for different sets of variables (i.e. thousand simulations). SC3 = sample campaign 3

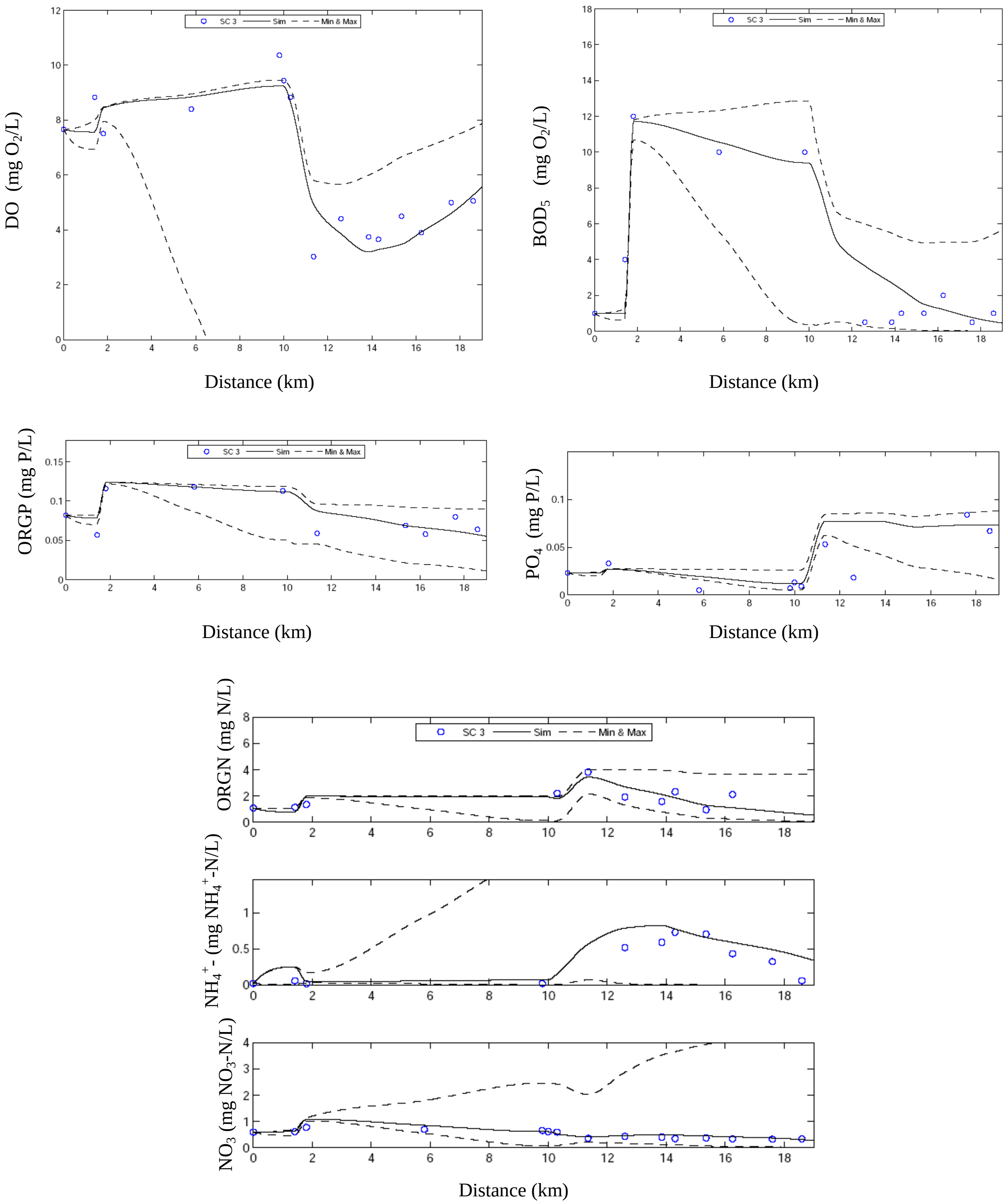


Fig. 5.3. Calibrated water quality model for dissolved oxygen (DO), five day biological oxygen demand (BOD_5), organic phosphorus (ORGP), phosphate (PO_4), organic nitrogen (ORGN), ammonia (NH_4^+) and nitrate (NO_3) in the Drava river. The actual simulation is given by the continuous line. The dotted lines indicate the maximal and minimal simulated values for different sets of variables (i.e. thousand simulations). SC3= sample campaign 3

A very wide range between the minimal and maximal simulated values of DO in the Drava river, in sampling sites located after the Čakovec lake can be appreciated in Fig 5.3 (abscissa 6 km). This phenomenon could be related with the reaeration processes simulated in this part of the river. The reaeration rate (k_a) for sampling sites located at the Čakovec lake were calibrated in the range between 0 and 2 1/d (Bowie et al., 1985), therefore, during the calibration process some of the 1000 simulation runs could take values of k_a close to zero (see Appendix E2.).

Model performance indicators (i.e. values of the determination coefficient, R^2) obtained during the calibration and validation processes for the physicochemical variables can be seen in Table 5.1. The evaluation of the calibration of the water quality model showed different values of model goodness of fit for each variable and each system modelled (i.e. South infiltration canal of lake Čakovec (C.L.) and the Drava river): (1) high values ($R^2 > 0.7$) for ORGP (only in the South infiltration canal C.L.), PO_4, NH_4^+, NO_3 and DO; (2) moderate values ($0.45 \leq R^2 \leq 0.7$) for BOD_5 in both systems and ORGP and ORGN in the Drava river; (3) low values ($R^2 < 0.45$) for ORGN in the South infiltration canal C.L. The validation of the water quality model showed: (1) high values ($R^2 > 0.7$) for ORGP, PO_4, ORGN, NH_4^+ and DO for the South infiltration canal C.L. and ORGP and BOD_5 for the Drava river; (2) moderate values ($0.45 \leq R^2 \leq 0.7$) for BOD_5 in the South infiltration canal C.L and NH_4^+, NO_3 and DO in the Drava river; (3) low values ($R^2 < 0.45$) for NO_3 in the South infiltration canal C.L. and PO_4 and ORGN in the Drava river.

5.3.3 Ecological river assessment model

5.3.3.1 Regression tree based on independent validation

Regression trees (RTs) were built to predict the EWQ (expressed as the MMIF index) as a function of physicochemical and hydromorphological river characteristics. The results of the implementation of the RTs based on an independent validation considering a dataset with and without outliers are presented in Table 5.2. The performance criteria for the RT indicate a moderate prediction capacity during the training process (monitoring campaign of September 2011) with: CCI = 50 %, RMSE = 0.73 and $r = 0.58$ for the complete dataset; CCI = 47 %, RMSE = 0.66 and $r = 0.56$ for the dataset without outliers and; $R^2 = 0.66$ for both datasets. During the validation process (monitoring campaigns of April and October

2010) the performance criteria indicated a low prediction capacity with: CCI = 40 %, RMSE = 1.64, R^2 = 0.20 for the complete dataset; CCI = 41 %, RMSE = 1.57, R^2 = 0.15 for the dataset without outliers and; r = 0.07 for both datasets. In general, the values of the performance indicators (CCI, RMSE, r and R^2) for the training and validation processes considering the complete dataset and the dataset with outliers deleted are similar, indicating that the deletion of the outliers does not increase the performance. A similar conclusion can be obtained when the training and validation datasets are combined to develop a regression tree (i.e. all dataset in Table 5.2) for the complete dataset (103 records) and the dataset without outliers (96 records).

Table 5.1. Average model performance indicators for the water quality model in the calibration and validation dataset (organic phosphorus (ORGP), phosphate (PO_4), organic nitrogen (ORGN), ammonia (NH_4^+), nitrate (NO_3), five-day biological oxygen demand (BOD_5) and dissolved oxygen (DO)).

***R^2* determination coefficient**		
South infiltration canal of Čakovec lake		
Variable	Calibration	Validation
ORGP	0.86	0.72
PO_4	0.88	0.95
ORGN	0.43	0.83
NH_4^+	0.95	0.92
NO_3	0.93	0.38
BOD_5	0.67	0.46
DO	0.90	0.79
Drava river		
Variable	Calibration	Validation
ORGP	0.62	0.71
PO_4	0.72	0.19
ORGN	0.55	0.24
NH_4^+	0.91	0.59
NO_3	0.95	0.52
BOD_5	0.44	0.74
DO	0.87	0.65

Table 5.2. Results of the regression tree using an independent validation method with and without outliers. (CCI: Correctly Classified Instances, RMSE: root mean square error, *r*: Pearson correlation coefficient, R^2: determination coefficient).

Dataset 1: Complete (with outliers)				
Dataset	CCI (%)	RMSE	*r*	R^2
Training	50	0.73	0.58	0.66
Validation	40	1.64	0.07	0.20
All	46	2.37	0.28	0.36

Dataset 2: Outliers deleted				
Dataset	CCI (%)	RMSE	*r*	R^2
Training	47	0.66	0.56	0.66
Validation	41	1.57	0.07	0.15
All	44	2.22	0.28	0.36

The RT trained and validated with the dataset without outliers is presented in Fig. 5.4a. This RT was assessed considering the ecological relevance and the practical use of the model. Four of the nine selected predictor variables were present in the RT (DO, average water depth, hydromorphological type and NH_4^+). A good EWQ (MMIF = 0.7) is defined by high concentrations of DO ($\geq$ 7.8 mg/L) and favourable hydromorphological conditions (Type = 1) such as natural bank structure, mixed bottom substrate, thin sludge layer, meandering, heterogeneous bank and bottom structure. Concentrations of DO higher or equal to 3.5 mg/L and lower than 7.8 mg/L and favourable hydromorphological conditions (Type = 1), leads to a moderate EWQ (MMIF = 0.51). Low concentrations of DO (< 3.5 mg/L) and moderately deep waters (average water depth $\geq$ 0.43 m) defines a bad EWQ (MMIF = 0.14). Poor EWQ is defined by low concentrations of DO (< 3.5 mg/L) and shallow waters (average water depth < 0.43 m). Moreover, unfavourable hydromorphological conditions (Type = 2) such as an artificial bank structure, tick sludge layer, straight waterway, homogeneous bank and bottom structure, together with concentrations of DO between 3.5 mg/L and 7.8 mg/L leads to Poor EWQ.

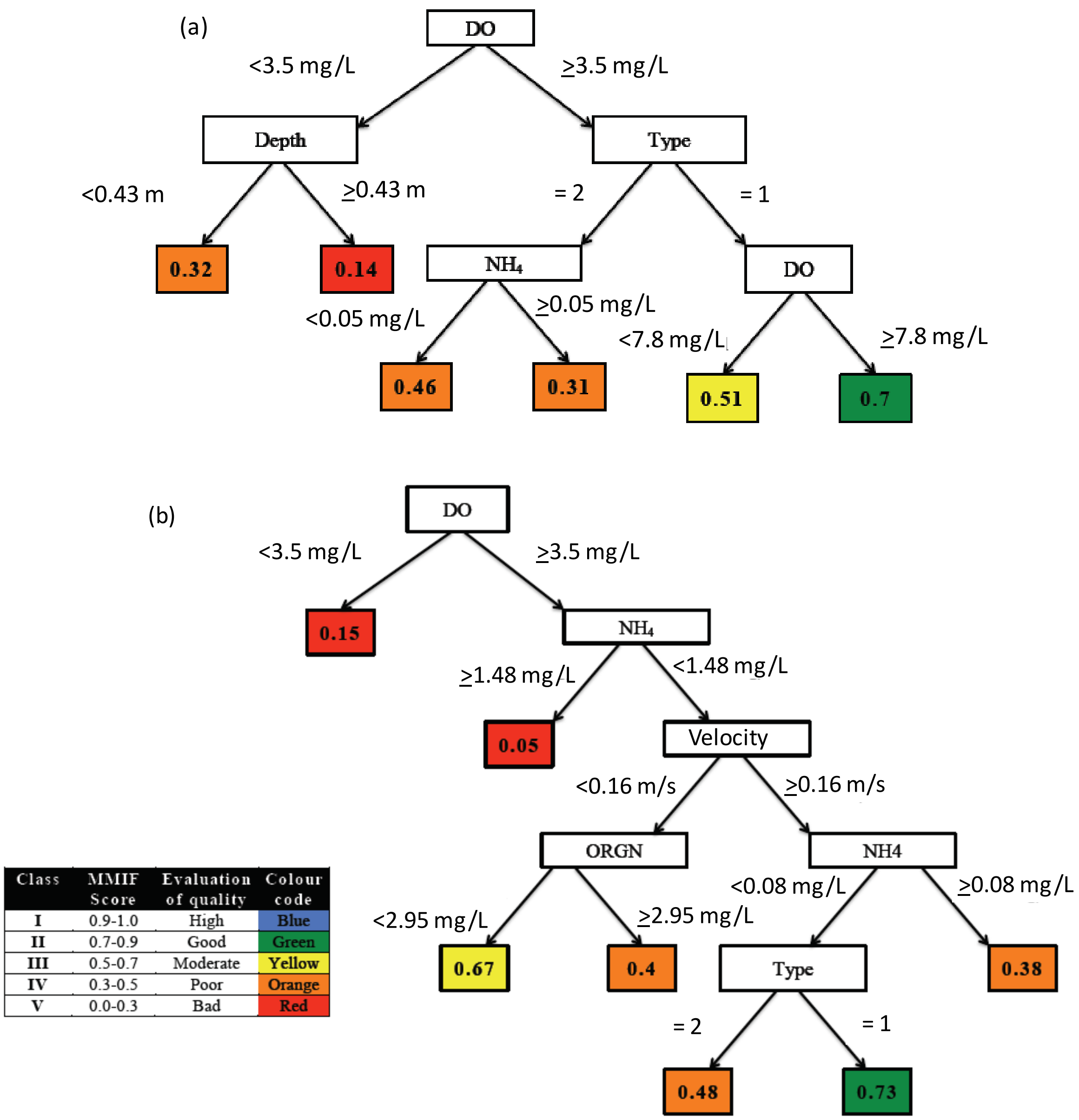


Fig. 5.4. Regression trees (RT) selected for predicting the EWQ based on the MMIF index. (a) RT built based on independent validation, (b) RT built based on internal validation, a resampling method based on a bootstrapping technique. (DO: dissolved oxygen, Type = 1: hydromorphological favourable, Type = 2: hydromorphological unfavourable, NH_4: ammonia, ORGN: organic nitrogen, depth = average water depth, velocity = average water velocity).

5.3.3.2 Regression tree based on internal validation

The results of the implementation of the RTs using internal validation with a bootstrapping technique and a stratification procedure for the dataset are presented in Table 5.3. Performance criteria for the 10 best models (based on CCI) of the 1000 bootstrap samples are shown in this table. The performance criteria for the best RT (stratified run 98) indicate a moderate prediction capacity (CCI = 59 %, RMSE = 1.94, r = 0.71 and R^2 = 0.44).

Table 5.3. Results of the regression tree using an internal validation (i.e. bootstrapping technique). Performance criteria for 10 best models of the 1000 bootstrap samples. (CCI: Correctly Classified Instances, RMSE: root mean square error, *r*: Pearson correlation coefficient; R^2: determination coefficient).

Stratified run	CCI (%)	RMSE	r	R^2
98	59	1.94	0.71	0.44
338	59	3.73	0.49	0.08
766	57	2.12	0.63	0.39
516	57	2.36	0.63	0.32
302	57	2.47	0.57	0.28
931	57	2.64	0.60	0.23
565	57	2.72	0.58	0.21
890	57	3.01	0.55	0.13
846	56	2.22	0.63	0.36
Average	45	3.57	0.47	0.04
Minimum	27	1.92	-0.01	1.22
Maximum	59	7.67	0.71	0.44
Variance	0.0	0.68	0.01	0.06

The RT obtained showed a significant ecological relevance (Fig. 5.4b). Four of the nine selected predictor variables were present in the RT (DO, NH_4^+, average water velocity, ORGN and hydromorphological type). A bad EWQ is defined by concentrations of DO lower than 3.5 mg/L (MMIF = 0.15) or DO concentrations higher or equal to 3.5 mg/L and NH_4^+ concentrations higher or equal to 1.48 mg/L (MMIF = 0.05). Good, moderate and poor EWQ are related to DO concentrations higher or equal to 3.5 mg/L, NH_4^+ concentrations lower than 1.48 mg/L and threshold values of average water velocity, ORGN and hydromorphological type respectively. Thus a good EWQ (MMIF = 0.73) can be obtained if there are favourable hydromorphological conditions (Type = 1), NH_4^+ concentrations lower than 0.08 mg/L and average water velocity higher or equal to 0.16 m/s. Concentrations of ORGN lower than 2.95 mg/L and an average water velocity lower than 0.16 m/s leads to a moderate EWQ (MMIF = 0.67). A poor EWQ is defined by concentrations of NH_4^+ higher than or equal to 0.08 mg/L and lower than 1.48 mg/L and average water velocity higher than or equal to 0.16 m/s (MMIF = 0.38). If there are unfavourable hydromorphological conditions (Type = 2), the NH_4^+ concentration lower than 0.08 mg/L and the average water velocity higher than or equal to 0.16 m/s a MMIF value of 0.48 (poor quality) can be obtained. Finally, average flow velocities below 0.16 m/s and concentrations of ORGN higher than or equal to 2.95 mg/L led to a MMIF value of 0.40 (poor quality).

5.3.4 Integrated ecological modelling and scenario assessment

The impact of upgrading the WWTP of the city of Varaždin with N and P removal on the EWQ of the south infiltration canal C.L. and the Drava river was evaluated based on three different wastewater treatment scenarios. Using the water quality and quantity models developed, it was possible to obtain profiles of average concentrations of DO, BOD_5, ORGP, PO_4, ORGN, NH_4^+, NO_3 and average values of water depth and water velocity for each scenario. This analysis allowed identifying that any change in the WWTP effluent quality has only an important effect on the water quality in the south infiltration canal C.L. (stations 5 to 9 in Fig. 5.1) and a marginal effect in the first section of the Drava river after the junction with the canal (station 10). As it was mentioned before, the south infiltration canal C.L. transports the discharge of the WWTP until the Drava river after lake Čakovec (between stations 8 and 10). The impact of the change in the discharge of the WWTP in all other sections of the Drava river located downstream of station 10 is practically negligible, due to the large dilution and long residence time effects.

Predictions of the EWQ for each scenario, expressed as the MMIF index value, were calculated using a RT and daily average predictions of the physicochemical and hydromorphological simulations as input variables. It was decided to use the RT based on internal validation because it showed a moderate prediction capacity compared with the RT based on independent validation, which had a very low prediction capacity during the validation process. An example of the EWQ predictions for each scenario in station 9, located at the end of the south infiltration canal C.L., is presented in Table 5.4. Here it is shown that the ecological impact of the three investment scenarios corroborates with our hypothesis. In order to improve the EWQ from bad to good in station 9 of the south infiltration canal C.L., it is necessary to upgrade the WWTP with N and P removal and the treatment of other point (e.g. the overflow of the WWTP) and diffuse pollution sources (i.e. scenario 3).

Table 5.4. Ecological water quality (EWQ) predicted for the three wastewater treatment scenarios considered in the integrated urban drainage system of the Drava river. Scenario 1: current situation; Scenario 2: upgrading of the wastewater treatment plant (WWTP) with nitrogen (N) and phosphorous (P) removal; Scenario 3: upstream treatment and upgrading of the WWTP with N and P removal. (MMIF: Multimetric Macroinvertebrate Index Flanders).

Section downstream of the WWTP discharge in the small channel (Section 9)		
		MMIF
Scenario		Class
1	Current situation	Bad
2	N and P removal	Poor
3	Upstream treatment plus N and P removal	Good

5.4 Discussion

5.4.1 Integrated ecological modelling framework

For the present study, the modular approach for model integration was adopted, where a model that simulates the outflow of a WWTP, a model for the river water quality and quantity with an ecological model were integrated. The IEMF implemented in this study evaluates the impact of different physicochemical and hydromorphological variables on the EWQ simultaneously. This modelling approach considers the hierarchy of these environmental variables at different scales (i.e. WWTP, river water quantity, physicochemical water quality and ecological assessment) in surface waters. The obtained IEMF can be used as a decision support tool for the evaluation of water management measures in order to improve the EWQ. The IEMF is data intensive, but it assists policy makers to take informed decisions regarding future investment programs for WWTP infrastructures.

The results of the implementation of the IEMF in the Drava river basin can be improved in some aspects: (1) The opinion of other stakeholders, such as the Croatian Electricity Company and the Croatian Environmental Agency should be included in the modelling process and scenarios building. In this study, mainly the sanitation company in Varaždin (VARKOM) participated. However, the expert knowledge and expertise of other stakeholders, could provide other investment scenarios that consider an increase in the minimum in-stream flow (‘environmental water requirement’) after the dams. By implementing these simulations it would be possible to evaluate scenarios that consider simultaneously the impact of upgrading the WWTP on the river ecosystem and the flow variations after the dam. (2) The ecological models implemented in the IEMF can be optimized by collecting more data with simultaneous measurements of physicochemical, hydromorphological and biological aspects. More samples should be collected, especially in surface waters characterized by a good and excellent ecological quality (i.e. to increase the stratified dataset). (3) The hydraulic and physicochemical river water quality modelling can be optimized by implementing complex hydraulic routing and complex pollutant transport routing methods (e.g. MIKE 11 model; DHI, 1999). By using these methods, assumptions such as uniform steady flow can be avoided, and, backwater and tidal effects can be considered. Moreover, rivers regulated by hydropower systems and dams, such as the Drava river, are affected by significant diurnal variation in flow, therefore, dynamic models are recommended. Additionally, urban and industrial effluents discharged on a batch basis or with significant variation in flow during different working shifts, together with low river flows could generate a peak of wastewater pollution with low dilution or low self-cleaning capacity of the river.

5.4.2 Ecological river assessment model

The regression tree developed in this study for predicting the MMIF index (Gabriels et al., 2010), was used to determine the relationship between a system's inputs and outputs using a training dataset. Two techniques were implemented for evaluating the predictive power of the RT in the IEMF (i.e. independent and internal validation). It was decided to use the RT based on internal validation, because it showed a better prediction capacity, it was statistically reliable, it was ecological relevant and it was applicable for decision support in water management. The resulting RT for predicting the MMIF showed that the most important environmental variables to assess and predict the EWQ were in order of importance DO, NH_4^+, average water velocity, ORGN and hydromorphological type. These results suggest that physicochemical indicators of organic pollution, such as low values of DO (< 3.5 mg/L), high values of NH_4^+ (≥ 1.5 mg/L) and high values of ORGN (≥ 3 mg/L), are the overriding factors to define bad, poor and moderate EWQ in the south infiltration canal of lake Čakovec and the Drava river. According to Chapman (1996), DO concentrations below 5 mg/L may adversely affect the functioning and survival of biological communities and below 2 mg/L may lead to the death of most fish. Concentrations of NH_4^+ in surface waters are typically less than 0.2 mg/L but may reach 2-3 mg/L. Higher concentrations could be an indication of organic pollution such as from domestic sewage, industrial waste and fertiliser run-off (Chapman, 1996). Similar results were reported by Pauwels et al. (2010) who applied regression trees to predict the MMIF in rivers in Flanders (Belgium) and reported that TP and DO were key variables to define moderate and good EWQ. Additionally, the RT showed the importance of driving forces such as the dam discharge and hydromorphological pressures. Thus, moderate values of average water velocity (≥ 0.16 m/s) and favourable hydromorphological conditions (Type=1) for the biological component (i.e. macroinvertebrates) together with concentrations of DO ≥ 3.5 mg/L and NH_4^+ < 0.08 mg/L leads to a good EWQ.

5.4.3 Integrated ecological modelling and scenario assessment

The implementation of different restoration options at the Drava river basin yielded three main results. First, there is a need for an integrated modelling approach that considers ecological aspects in the water management of this river. Second, any change in the WWTP effluent quality has an important effect only to the water quality in the south infiltration canal of lake Čakovec in which it is discharging and has a marginal effect in the first section of the Drava river after the junction with the canal. Downstream from the WWTP effluent discharge point, a better effluent quality did not have a significant impact on the river ecological water quality due to dilution and long residence time effect. Third, in order to change the EWQ from bad to good state in station 9 of the south infiltration canal C.L., it is necessary to upgrade the WWTP with N and P removal and to provide the treatment of other point (e.g. the overflow of the WWTP) and diffuse pollution sources (i.e. scenario 3).

Therefore, additional pollution sources present in the study area should be monitored and remediated. Non-point sources of pollution are assumed to have a greater relative importance in water quality management as point sources have come under increasingly stringent control. Unfortunately, non-point source loads are often driven by rainfall events and thus both the wasteload and flow vary significantly over time (Reichert et al., 2001b). Among the most important causes of acute pollution are combined sewer overflows (CSOs), especially considering the DO concentrations (Hvitved-Jacobsen, 1982). CSOs of WWTP are generated in sewer systems, in which sewage and runoff, from the catchment area are transported to the WWTP for purification and subsequent release into the receiving water. However, when the amount of runoff exceeds the given hydraulic capacity of the plant, (diluted) wastewater is discharged to the receiving river directly, which can be seen conceptually as a bypass of the WWTP. The degradation of physicochemical and biological quality of urban receiving waters by discharges from CSO and surface water outfalls (SWO's) has been documented (Hvitved-Jacobsen, 1982; Mullis et al., 1997). CSOs can be dangerous for the ecosystem as well, regarding physicochemical or combined factors at different time scales due to shear stress, non-ionized ammonia, oxygen depletion, (sedimentation of) suspended solids, persistent organic substances, metals, nutrients, among others (Borchart and Sperling, 1997). Therefore, usually rain storm tanks are considered useful to minimize the consequences of CSO's in several respects, and may be located either in the sewer network or at the treatment facility. Many studies tend to control CSO's and SWO's by using these facilities (Bwalya, 1996; Breur et al., 1997). Such type of facilities could be implemented in the WWTP to reduce the impact of wastewaters generated in the city of Varaždin, Croatia. Moreover, a redesign and reconstruction of the WWTP, with an extra unit for more wastewater treatment capacity, is required to avoid an overload of the secondary clarifiers.

5.5 Conclusions

The proposed model integration between the WWTP, water quality, water quantity and river ecological assessment models is a suitable approach to evaluate the impact of sanitation infrastructures, such as WWTPs, on the ecological state of the receiving river. The IEMF was used as a tool to develop a model that integrated physicochemical, hydromorphological and ecological aspects in the water management of the Drava river. Yet, the shortcomings of this approach are acknowledged; it is data intense as it requires WWTP, water quantity and quality models for a specific river, and extended (a)biotic data.

Chapter 6: General discussion and conclusions

The overall aim of this study was to develop and to evaluate an integrated ecological modelling framework for decision support in river management. To this end, a conceptual framework for integrated modelling called IEMF (*Integrated Ecological Modelling Framework*) was developed (presented in Chapter 1) and tested in three case studies (Chapter 3-5). The IEMF combined the results and information obtained from field data and integrated river water quality and quantity models with aquatic ecological models based on data-driven modelling techniques. By following the IEMF the link between physicochemical and hydromorphological pressures with the ecological state of the river can be established. This generic modelling framework can be used for decision support in river management and water policy as it allows simulation analysis to assess different river management options.

6.1 Integrated ecological river modelling framework proposed

Up to now, several conceptual (modelling) frameworks, developed as decision support tools in river water management, do not consider the simultaneous effect of hydromorphological disturbances and physicochemical pollution on the river ecology. The DPCER (Rekolainen et al., 2003) and SPEAR (von der Ohe et al., 2009) frameworks consider chemical and ecological states of the receiving river, whereas the PHABSIM (Bovee et al., 1998; USGS, 2001) and the DPSI (Jähnig et al., 2012) consider hydromorphological / hydraulic and ecological states. Therefore, this research aimed to develop the IEMF that covers the gaps that other conceptual frameworks have until now. This framework considers physicochemical pressures, such as the discharge of wastewater treatment plants (WWTP) and hydromorphological pressures, such as changes in water course, current velocity, water depth, riverbed sediment composition and bank structure. Such comprehensive evaluation could not be achieved when looking at each individual component of the system separately (i.e. sewer system, WWTP, dam and receiving river).

The implementation of the IEMF with three case studies, allowed identifying that there is a general need for an integrated ecological modelling approach in the water management of the three evaluated rivers (i.e. Cauca river in Colombia, Cuenca river in Ecuador and

Drava river in Croatia). The results show that the integration of ecological models (e.g. habitat suitability and river ecological assessment) in hydraulic and physicochemical water quality models (e.g. MIKE 11, QUAL2Kw and RWQM1) has an added value for decision support in river management and water policy. The IEMF assists the water quality managers (authorities) in the following topics: (1) through the conceptual elements considered in the IEMF (driving forces, pressures, physicochemical, hydromorphological and ecological state and response) a better interpretation of the ecological river state can be possible, the causes of the state of a river can be detected and assessment methods can be optimised; (2) the IEMF can allow for calculating the effect of future investments in sanitation infrastructures (e.g. collection and treatment of wastewater) and river restoration actions on aquatic ecosystems and supporting the selection of the most sustainable options; (3) the IEMF can allow for predicting and assessing the achievement of predefined ecological water quality objectives. These objectives can be represented by threshold values of ecological indices (e.g. BMWP, IBIAP, MMIF) or by the improvement of habitat conditions for targeted aquatic species (e.g. pollution sensitive macroinvertebrates) and; (4) the IEMF can help to find the major gaps in our knowledge of river systems and help to set-up cost effective monitoring programmes. The integration of models is a key aspect for environmental decision making.

The novelty and technical advance of the IEMF in the integration of models towards the assessment of the ecological state of rivers have been demonstrated in the three case studies: (1) simultaneous assessment of the impact of hydromorphological pressures and physicochemical pollution on the ecological river water quality; (2) the use of different approaches for water quantity and quality modelling (i.e. dynamic and steady state) with detailed and specific models (e.g. detailed physical habitat and WWTP processes) which can be integrated with aquatic ecological models; (3) development of ecological models based on specific characteristics of the studied river; (4) flexibility for updating or replacing the (ecological) models by better models when available, without having to change the framework. This demonstrates the flexibility, applicability and transferability of the IEMF to other regions in the world.

The validation of the IEMF was performed in terms of its applicability, as decision support tool in river water management, in three rivers with different geographical locations, altitude, size and pollutions problems. Thus, two deep lowland rivers located in a tropical region (Cauca river in Colombia, Chapter 3) and a temperate zone (Drava river in Croatia, Chapter 4) and one shallow mountain river in a tropical region (Cuenca river in Croatia, Chapter 5) were evaluated. This analysis allowed identifying that: (1) different types of water quantity and quality models (dynamic or steady state) could be required according to the level of model complexity considered (see further analysis in section 6.2.2); (2) the selection of the type of data-driven modelling technique for the ecological models depends on the type of data (dichotomous (presence/absence), count data or continuous data) and availability of data. Thus, in this research a threshold value of 30 records with simultaneous measurements of physicochemical, hydraulic/hydromorphological and biological information was considered to choose between parametric methods such as GLM (e.g. LRM and NBRM) or non-parametric methods such as decision tree methods (CT, RT and MT) (see further analysis in section 6.2.3).

6.2 Practical recommendations for integrated ecological modelling of rivers

This PhD study consisted of four major activities: (1) integrated data collection; (2) model implementation for hydraulic and physicochemical water quality models and wastewater treatment plant processes; (3) model implementation for ecological models and; (4) the integration of models to support decision making in river management. The aim of this section is to link the results and discussions in the previous chapters and present some general and practical recommendations with regard to the development and application of integrated ecological modelling of rivers for decision support in water management.

6.2.1 Integrated data collection

The main findings regarding integrated data collection obtained in this research are:

1. The main limitation of the IEMF is the availability of physicochemical, hydromorphological, hydraulic and biological data that are collected simultaneously. Therefore, a change in the river monitoring strategy towards collection of data which include simultaneous measurements of these variables is required.
2. In general the data used in this research (in the three case studies) was lacking sufficient records with excellent and good ecological water quality (EWQ), consequently these EWQ classes were unequally represented in the dataset, especially

in the case study of Croatia. This situation is related with the impact of pressures, such as physicochemical pollution and hydromorphological disturbances, on the EWQ. In this case, it is necessary to search for extra sampling locations located upstream of the sampling points influenced by these impacts. This effort was performed in the third monitoring camping of the Drava river in Croatia, however sampling locations with excellent and good EWQ were limited.

3. In this case, a stratification procedure for the dataset (Everaert et al., 2013), in order to guarantee the same number of instances for each EWQ class represented in the set and to avoid biased models, can be implemented.
4. The data available for building the ecological models, need pre-processing before it can be used for the coupling of models. This data needs a good and accurate analysis, in order to identify: (1) possible outliers; (2) collinearity between predictor variables and; (3) relationships between the response variables (i.e ecological indices or presence/absence of macroinvertebrates) and the predictor variables. It was found that graphical statistical tools such as box plots and Cleveland dot plots (Zuur et al., 2010) help to evaluate possible outliers. Moreover, preliminary mass balance analysis with the water quality models help to evaluate the reliability of the data and to identify possible outliers. Collinearity assessment can be performed by a Principal Component Analysis (PCA) with a varimax rotation, to maximise the independence of the Principal Components and the Spearman rank (S) or the non-parametric correlation coefficient Kendall's (τ). These two correlation coefficients are better suited for this analysis compared with the Pearson correlation coefficient, because the S coefficient makes no assumptions about linearity in the relationship between the variables (Zuur et al., 2009) and the τ coefficient can deal better with outliers and extreme distributions of the variables (Willems et al., 2008). Relationships between the response and the predictor variables can be assessed by using the S or τ correlation coefficients.

6.2.2 Hydraulic and physicochemical water quality models implementation

As Shanahan et al. (2001) properly pointed out, the construction of a river water quality model must be based on the logical development of the elements in the model, which can vary with local conditions. These authors indicated that the details of, especially the more complex, models and choice of algorithms vary with the type of information available, the complexity of the system and the environmental problem assessed. For example, for shallow mountain rivers, such as the Cuenca river in Ecuador, steady state models such as the QUAL2Kw are well suited, if calibration and validation monitoring campaigns are taken following a water-mass volume from the source to the mouth (see section 4.2.3.1 in Chapter 4.). Thus, variation in flow and water quality conditions can be monitored and simulated in a water-mass volume from the source to the mouth. On the other hand, for deep lowland rivers such as the Cauca river in Colombia (see section 3.2.4.1 in Chapter 3.) dynamic models, such as the MIKE 11 are more appropriate. This aspect is especially important for future research in the Drava river in Croatia, which is a deep lowland river, however, in this study the dynamic flow conditions were simplified by using a surrogate model, such as the conceptual hydraulic model based on the CSTRS approach (see section 5.2.3.2 in Chapter 5.). By using dynamic models, assumptions such as uniform steady flow can be avoided, and, backwater and tidal effects can be considered. These conditions are particularly important in regulated rivers, by hydropower systems and dams, which are affected by significant diurnal variation in stream flow. Moreover, the hourly fluctuations of pollution load from urban and industrial wastewater discharges can be considered in dynamic models.

6.2.3 Ecological model implementation

The main findings regarding the ecological model implementation obtained in this research are:

1. The approach followed in this study for the ecological modelling was to use RT, MT and GLM techniques. Typically, integrated data collection required for the IEMF is prone to generate small datasets ($n < 30$). The datasets in the case study in Colombia had 15 records, whereas in Ecuador and Croatia 60 and 96 records were available respectively. In case of small datasets, parametric methods such as GLM, which are generally more efficient on small datasets than non-parametric methods such as decision tree methods (Vayssières et al., 2000) are recommended. However, when using GLM methods such as logistic regression, Poisson, quasi-Poisson or negative binomial regression, it is necessary to validate the regression technique implemented. Statistical stools, such as diagnostic plots for model adequacy and the lack-of-fit test

available in the package CAR in the software R (Fox and Weisberg, 2011) are well suited for this purpose.

2. Two types of ecological models based on data-driven techniques were developed in this research: (1) river ecological assessment and; (2) species distribution models to predict the habitat suitability for selected species of macroinvertebrates. These models are useful tools for predicting changes in river networks due to disturbances or restoration efforts. In addition, these models are a fast and effective way to predict EWQ deteriorations or improvements in river systems and allow users to deduce information about a river system that is sometimes unfeasible and very time consuming to monitor. In the three case studies (i.e. rivers in Colombia, Ecuador and Croatia), there is a need for the development of practical modelling tools providing accurate ecological assessment of rivers and species conditions. These modelling tools could include aquatic habitat suitability models such as LRM, and ecological assessment models such as NBRM, MT and RT. This should allow preserving habitats and species, stop degradation and restore water quality. Insight in the habitat preferences of aquatic organisms will be helpful to our river restoration management plans and vision building. An understanding of the causal mechanisms and processes that affect the ecological water quality and shape macroinvertebrate communities at a local scale has important implications for conservation management and river restoration. Habitat suitability models have received criticisms both for being too complicated or too simplistic; one of the key issues has been the development and transferability of the preference relationships. However, as few alternatives are available, they remain key tools for environmental quality assessment (REBECCA, 2004).
3. The evaluation of the predictive performance and robustness of the ecological models is a vital step in model development. Such assessment serves three main purposes: (1) it allows determining the suitability of a model for specific applications; (2) it provides a basis for comparing different modelling techniques and competing models and; (3) it allows identifying aspects of a model most in need of improvement (Pearce & Ferrier, 2000). However, apart from the statistical reliability also the applicability and the ecological relevance are important aspects for model selection (Larocque et al., 2011; Everaert et al., 2012). In an applied sense, models have their greatest utility when they can be used predictively and not simply as a means of exploring relationships in a dataset (Rushton et al., 2004). The GLM and decision trees implemented were assessed in three steps. In a first step, the models were evaluated based on mathematical criteria,

such as percentage of Correctly Classified Instances (CCI), the Cohen's kappa coefficient (K) and the area under the receiver-operating-characteristic (ROC) curve called AUC. Secondly, all statistically significant models were verified based on ecological insight. It was found that including stakeholders in the model building process can improve the model reliability. This was shown in the case study of the Drava river in Croatia, with the participation of the sanitation company in Varaždin (VARKOM) and the developers of the ecological models (section 5.4.1 in Chapter 5.). Stakeholders evaluated all statistically reliable data-driven models for their ecological relevance. In case that, biological inconsistencies were found, these models were discarded. This is similar to previous studies, such as the reported by Voinov and Bousquet (2010), who concluded that it is vital to include stakeholders during the modelling process. A third criterion to evaluate the ecological models was by verifying the applicability and practical use for decision support in water management.

4. It was found that the foreseen investments in sanitation infrastructure and current river restoration programs considered for the river basins in the three case studies are not enough to provide a good ecological water quality. Advanced investments, such as the collection and treatment of all domestic and industrial wastewater received by the rivers, the control and monitoring of the diffuse pollution sources and the upgrading of the existing WWTP, with nitrogen and phosphorous removal are required. Moreover, it was identified that combined sewer overflows in the Drava river in Croatia, generate a high negative impact on the ecological river water quality. Similar findings, concerning physicochemical and biological quality of urban receiving waters were reported by Hvitved-Jacobsen (1982) and Mullis et al. (1997).

6.3 Integrated ecological modelling with stakeholders

The optimization of freshwater ecosystem services and the sustainability of water resources depend on the participation of stakeholders during the modelling and decision taking processes (Molle, 2009; Voinov and Bousquet, 2010; Vanrolleghem, 2010a). Additionally, when multiple impacts, such as habitat degradation and water pollution, affect (simultaneously) the ecological water quality, decision support tools such as the IEMF are required. This type of integrated modelling framework allows determining an optimal balance between the different stakeholder activities in the integrated water resources management. Thus, the expert knowledge and expertise of the different stakeholders in the river basin can be included during the implementation of models and the setting-up of simulations scenarios for water quality management and pollution control.

O'Kane (2008) incorporated the term "social calibration" when models are used in real-life-decision-making frameworks and education. This author suggested involving stakeholders with the best knowledge of the aquatic system in question, rather than purely numerical calibration without an insight. In model calibration and validation, every mismatch between a prediction and a measurement raises the question, why? Mismatches can emerge from errors in the model, errors in the data, or errors in both the model and the data (O'Kane, 2008). Answering such questions improves the model. These stakeholders are shown animated graphical output from the model for historical events and asked if they are true. When the answer is yes, this step builds credibility and acceptance of the model (O'Kane, 2008). Only then, the model can be used to examine engineering alternatives (e.g. water quality management and pollution control scenarios) that affect stakeholders.

By providing the IEMF, the integration of different models, data, information resources and stakeholders knowledge can be performed. In this PhD study several stakeholders of river basins were involved (e.g. environmental authorities, municipalities, sanitation companies and industries) however, the opinion of other stakeholders, such as hydropower companies and farmers, among others, is also recommendable. By becoming more aware of the needs of the stakeholders as policy makers and their operating constraints, models were developed target policy relevant issues by integrating (ecological) specific norms or indicators. This analysis is in concordance with what Vanrolleghem (2010a) stated about improving stakeholder involvement with participatory modelling in decision-making processes.

6.4 Recommendations for further research

There are other studies that can be developed in the future and can contribute to the integrated ecological modelling as decision support tools in river management. These topics are discussed and suggestions how such research can be set up are presented.

- **Data availability and accessibility in view of integrated river water quality assessment and model-based water management:**

The application of models in ecology is almost compulsory if we want to understand the function of such a complex system as an ecosystem (Jorgensen and Bendoricchio, 2001). However, the knowledge of ecological processes in ecosystems and the information available for a very deep insight of these processes have been much less developed and accessible compared with other science fields such as hydrodynamic or hydromorphological and physicochemical processes. Thus, the use of (predictive) ecological models might result in a more rational analysis of aquatic ecosystems and help to develop and to improve river assessment systems.

Today river water quality assessment is mainly based on discrete monitoring campaigns, with time intervals of several hours, weeks, months or even years. For the study of highly dynamical processes such sampling schemes are often insufficient to make a reliable assessment of the river status. In those cases, the application of automated measurement stations for continuous water quality monitoring together with the study of biological indicator species, such as macroinvertebrates, are complementary tools for river quality assessment. Considering the seasonality of the life cycles of some macroinvertebrates (e.g. insects), it is recommended to perform seasonal monitoring campaigns, at least two times per year, one in dry season and other in rainy season (spring or summer and autumn). Having relatively long life cycles and being confined for most part of their life to one locality on the river bed, aquatic macroinvertebrates act as continuous monitors, integrating water quality over a longer period of time (weeks, months, years). Biological indicator species are unique environmental indicators as they offer a signal of the biological condition in a watershed.

Mechanistic and data-driven models are clearly affected by the type of variables that are collected, thus before modelling processes, it is necessary to define what type of variables need to be collected in the field. Therefore, model development studies need to be based on questions from (water) managers. Once these are identified and the necessary models and variables are known, a relevant data acquisition has to be set up (Goethals, 2005). Guisan and Zimmerman (2000) properly pointed out that too many static modelling exercises are still based on field data from observational studies, lacking a sampling design strategy.

- **Linking ecological models to social-economic models and stakeholder information needs:**

There is a strong need for combining ecological tools with social-economic valuation methodologies to develop insight in the economic benefits of the goods and services supplied by the terrestrial and freshwater ecosystems. The WFD provides an integrated approach to catchment management that, while widely accepted, is characterized by scientific, socio-economic and administrative complexities. In this context, expert knowledge-based models such as fuzzy models or Bayesian Belief Networks can be utilised to provide synthesis of these complex processes and to identify the likely response within and among domains of natural and anthropogenic changes (e.g. Mouton et al., 2009; Landuyt, et al., in press). As Irvine et al. (2002) clearly stated it is difficult to envisage cost-effective and meaningful management without such aids.

Environmental decision-making is extremely complex due to the intricacy of the systems considered and the competing interests of multiple stakeholders. Therefore, the expert knowledge and expertise of the different stakeholders in the river basin should be included during modelling and decision taking processes. Consequently, the development and application of decision support tools, such as integrated ecological modelling in river water management are necessary. This integrated approach serves, besides its function as a decision support tool, as a communication tool for providing information to the river managers. This approach tries to break the paradigm of decision makers that often complain that environmental models are not readily available, accessible or understandable (Liu et al., 2008). However, in order to make this modelling framework readily understandable for decision makers, the creation of a friendly user interface would be beneficial.

- **Linking ecological models to climate, land-use, hydrological, river water quality and quantity and other physical models:**

The use of mathematical models within an integrated river water quality management requires a transcendence of scales and disciplines. Traditionally, scientists develop and run models within well-defined domains of applicability, and the need for integration of scales and, particularly, disciplines can restrict model use. Models that have been developed at the mesoscale or microscale level, where the impact of physical habitat changes on river biology occurs, cannot be automatically applied at the river basin scale without serious consideration of suitability, robustness and reliability. Therefore, there is increasing recognition that heterogeneous catchments require a range of modelling approaches, with data collections made at appropriate spatial and temporal scales (Irvine et al., 2002).

Examples of the application of integrated ecological modelling as decision support tools for supporting the implementation of the WFD goals at river basin scale are presented by Vandenberghe et al. (2005) and van Griensven et al. (2006). These authors coupled the Soil and Water Assessment Tool (SWAT) results to ecological models. Combined impacts of nutrient inflows from agricultural and wastewater discharges from industries and households and habitat modifications due to human disturbances, can be assessed with water quality models developed in ESWAT, a SWAT2000 version that was extended with hourly hydrological and water quality processes (van Griensven and Bauwens, 2001). SWAT is an open-source software which has high level of flexibility for a wide range of applications by allowing the users to do case-specific adaptation to the source code and linking it to other models and modelling tools (van Griensven et al., 2006). SWAT is a conceptual model that operates on a daily time step, functions on the catchment scale and includes processes for the assessment of point and complex diffuse pollution sources. Model subbasin components can be divided as follows: hydrology, weather, sedimentation, soil temperature, crop growth, nutrients, pesticides and agricultural management. Thus, further linkage of SWAT to ecological assessment tools, land use prediction tools and climate change models, shows that SWAT can play an important role in integrated ecological modelling as decision support tools in river management.

- **Model uncertainty in the IEMF:**

The integration of different data, several information sources and different models in one framework such as the IEMF, has its downsides, such as uncertainty propagation in the integrated model. The propagation of the error coming from the water quality and quantity models to the ecological models can induce a rather large error in the output. However, this error propagation was not considered in this study, but should be assessed in future studies. Cluckie and Xuan (2008) addressed the issue of uncertainty propagation in an integrated model for rainfall prediction systems used for operational real-time flood forecasting. This type of analysis can be implemented in the IEMF, focusing on multi-model inference techniques based on the information-theoretic approach and model averaging.

Appendices

Appendix A. Detailed description of Materials and methods

Appendix A.1 Generalized Linear Models

The approach followed for building some of the ecological models used in the case studies of the Cauca river (Colombia; Chapter 3.) and Cuenca river (Ecuador; Chapter 4.) was to use multivariate statistics based on Generalized Linear Models (GLM). It was decided to implement two GLM techniques, a logistic regression model (LRM) for predicting occurrence of macroinvertebrates (for both case studies in Colombia and Ecuador; Chapters 3. and 4.) and a negative binomial regression model (NBRM) for predicting the value of the BMWP-Colombia (only for the Cauca river; Chapter 3).

LRM estimates the probability of a response variable (presence/absence) given a set of explanatory environmental variables (e.g. DO, BOD_5). LRM can be used for the simultaneous analysis continuous variables such as water depth and water velocity and categorical variables (e.g. substrate type). Based on the presence/absence data, a response curve of a species describes the probability of the species being present, *p*, as a function of environmental variables. The response variable is transformed by the logit link function, which transforms bounded probabilities (between 0 and 1) to unbounded values (Ahmadi-Nedushan et al., 2006). The LRM is expressed as:

$$g_{(X)} = Logit(p) = \ln\left[\frac{p_i}{1-p_i}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m \qquad \text{(A.1)}$$

$$p = \frac{\exp(g_{(x)})}{1+\exp(g_{(x)})} = \frac{1}{1+\exp(-g_{(x)})} \qquad \text{(A.2)}$$

$$p = \frac{1}{1+\exp(-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m))} \qquad \text{(A.3)}$$

where $g_{(x)}$ is the linear combination of environmental factors; p_i is the probability that a species is present in a cell or the probability that a habitat cell would be suitable for a species; β_0 and $\beta = \{\beta_1, \ldots \beta_m\}$ are regression constants and $X = (X_1, \ldots, X_m)$ is a vector of *m* predictor variables.

GLM techniques such as Poisson, quasi-Poisson or negative binomial regression are the most common approaches used for predicting count data (which are a non-negative integer values). Examples of count data are trend: (1) abundance data (e.g. abundance of species); (2) density (numbers (which are counts!) per volume (or area, depth range, etc)) and; (3) (ecological) indices (e.g. BMWP-Colombia, which ranges between 0 and 120). These types of data cannot be represented by Gaussian distribution, therefore other type of distributions such as the Poisson distribution or the negative binomial distribution should be evaluated trend (Zuur et al., 2009). If there is overdispersion in the data (dispersion parameter (Φ) higher than one), then the quasi-Poisson or the negative binomial regression can be used for modelling. If there is low overdispersion the quasi-Poisson regression is the option, otherwise, negative binomial regression can be used. The selection of the suitable GLM model (i.e. Poisson, quasi-Poisson or negative binomial regression), can be evaluated by plotting the residuals after fitting the regression models (Zuur et al., 2009). The NBRM is expressed as:

$$E(Y) = \mu \qquad \mathrm{var}(Y) = \mu + \frac{\mu^2}{\phi} \tag{A.4}$$

$$g_{(X)} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m \tag{A.5}$$

$$\mu_{(X)} = \exp(g_{(X)}) = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m) \tag{A.6}$$

where E (Y) is the expected value of Y, var(Y) is the variance of Y, Φ is the dispersion parameter.

Appendix A.2 Predictive performance criteria in LRM:

Current practice in species distribution modelling suggests applying at least two different performance criteria for model evaluation. In general, there are two approaches to assess the model performance of LRM: (1) threshold-dependent approaches, such as the percentage of Correctly Classified Instances (CCI) and Cohen's kappa coefficient (Cohen's *K*) and; (2) threshold-independent approaches, such as the area under the receiver-operating-characteristic (ROC) curve called AUC. The confusion matrix (Table A1) is the basis to calculate CCI and *K* by following the equations:

$$CCI = \frac{a+d}{n} \tag{A.7}$$

$$\text{Cohen's } K = \frac{((a+d)/n) - ((a+b)(a+c) + (c+d)(d+b)/n^2)}{1 - ((a+b)(a+c) + (c+d)(d+b)/n^2)} \tag{A.8}$$

where a is the number of true positives; b the number of false positives; c the number of false negatives; d the number of true negatives and n the total number of instances.

Table A.1 The confusion matrix as a basis for the performance measures with true positive values (TP), false positives (FP), false negatives (FN) and true negative values (TN).

		Observed	
		+	-
Predicted	+	a (TP)	b (FP)
	-	c (FN)	d (TN)

Gabriels et al. (2007) suggest the following ranks of model performance for Cohen's *K* values in a freshwater ecological context: 0.0–0.2: poor; 0.2–0.4: fair; 0.4–0.6: moderate; 0.6–0.8: substantial; and 0.8–1.0: excellent. The AUC, which ranges from 0.5 to 1.0, gives an idea of the discrimination capacity of the model. A model with good discrimination ability is the one that can correctly discriminate between occupied (presence) and unoccupied (absence) sites in an evaluation dataset. Hosmer and Lemeshow (2000) and Pearce and Ferrier (2000) suggest that for a model with perfect discrimination the AUC=1 and for a model with no discrimination ability the AUC=0.5. Values between 0.5 and 0.7 indicate poor discrimination capacity, values between 0.7 and 0.9 indicate reasonable discrimination ability appropriate for many applications and rates higher that 0.9 indicate a very good discrimination.

Appendix B. Summary of data used in the three case studies in this research

Table B.1 Average, minimum and maximum values of the assessed environmental variables in the Cauca River in Colombia, based on 15 samples during the period 1996-2005. SD: Standard Deviation; DO: dissolved oxygen; BOD_5: five-day biological oxygen demand; BMWP-Colombia (Zúñiga and Cardona, 2009). Target macroinvertebrate taxa selected: Ephemeroptera (pollution sensitive taxon) and Haplotaxida (pollution tolerant taxon).

Variable	Unit	Mean	Minimum	Maximum	SD
Ecological predictor variables					
DO	mg O_2/L	4.17	0.3	6.89	2.2
Temperature	C	22.7	18.0	26.4	2.4
BOD_5	mg O_2/L	4.10	0.12	15.45	4.45
Flow	m^3/s	218.6	83.3	509.0	131.3
Water depth	m	4.4	2.1	7.2	1.6
Water velocity	m/s	0.7	0.5	1.2	0.2
Ecological response variable					
BMWP-Colombia	dimensionless	28	3	55	17
Ephemeroptera	present (n=6)				
	absent (n=9)				
Haplotaxida	present (n=9)				
	absent (n=6)				

Table B.2 Average, minimum and maximum values of the assessed environmental variables in the Cuenca River in Ecuador. Observed characteristics in the Tarqui, Yanuncay, Machangara, Tomebamba and Cuenca rivers, based on 60 samples during the period 1997-2009. SD: Standard Deviation; DO: dissolved oxygen; BOD_5: five-day biological oxygen demand; FC: Faecal Coliforms using the most probable number (MPN) method; IBIAP: Biotic Integrity Index using aquatic macroinvertebrates (Carrasco, 2008). Target macroinvertebrate taxa selected: Trichoptera (pollution sensitive taxon), and Physidae (pollution tolerant taxon).

Variable	Unit	Mean	Minimum	Maximum	SD
Ecological predictor variables					
DO saturation	%	7.3	2.2	8.9	1.3
Temperature	C	15.9	10.0	21.2	3.0
BOD_5	mg O_2/L	8.0	0.4	103.0	19.4
FC	MPN/100mL	2.8×10^5	1.7×10^1	7.9×10^6	1.0×10^6
Flow	m^3/s	6.5	0.1	41.1	7.7
Water depth	m	0.6	0.1	1.2	0.3
Water velocity	m/s	0.7	0.2	2.1	0.3
Ecological response variable					
IBIAP	dimensionless	7	4	15	2
Trichoptera	present (n=23)				
	absent (n=37)				
Physidae	present (n=42)				
	absent (n=18)				

Table B.3 Average, minimum and maximum values of the assessed environmental variables in the Drava River in Croatia, based on 103 samples during the period 2010-2011. SD: Standard Deviation; DO: Dissolved oxygen; BOD_5: five-day biological oxygen demand; ORGN: organic nitrogen; NH_4^+: ammonium; NO_3: nitrate; PO_4: phosphate; ORGP: organic phosphorus; MMIF: Multimetric Macroinvertebrate Index Flanders (Gabriels et al., 2010).

Variable	Unit	Mean	Minimum	Maximum	SD
Ecological predictor variables					
DO	mg O_2/L	5.6	0.5	12.7	2.5
BOD_5	mg O_2/L	4.3	0.5	35.0	5.7
ORGN	mg N/L	1.85	0.08	6.31	1.29
NH_4	mg N/L	0.33	0.002	3.07	0.51
NO_3	mg N/L	0.56	0.04	1.81	0.33
PO_4	mg P/L	0.11	0.002	2.27	0.25
ORGP	mg P/L	0.1	0.002	0.85	0.12
Water depth	m	1.92	0.12	10.0	2.92
Water velocity	m/s	0.35	0.002	1.03	0.29
Ecological response variable					
MMIF	dimensionless	0.41	0.05	0.85	0.19

Appendix C. Supplementary information of the case study of Colombia (Chapter 3.)

Appendix C1. Dataset pre-processing in the case study of the Cauca river

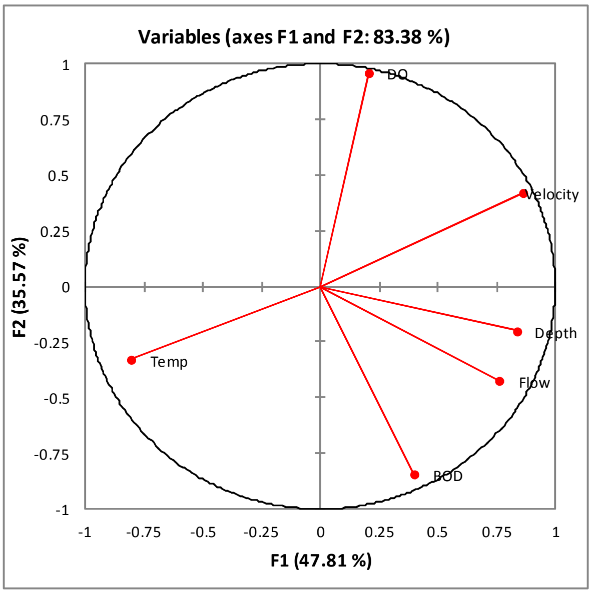
(a) Principal Component Analysis (PCA) with a varimax rotation

Eigenvalues:

	F1	F2	F3	F4	F5	F6
Eigenvalue	2.868	2.134	0.453	0.341	0.157	0.047
Variability (%)	47.807	35.571	7.545	5.678	2.620	0.779
Cumulative %	47.81	**83.38**	90.92	96.60	99.22	100.00

Factor loadings:

	F1	F2	F3	F4	F5	F6
Flow	**0.759**	-0.423	-0.112	0.480	0.035	0.011
Depth	**0.836**	-0.201	0.473	-0.083	-0.165	0.048
Velocity	**0.861**	0.421	0.107	-0.098	0.211	-0.122
Temp	**-0.806**	-0.328	0.424	0.146	0.202	0.009
BOD	0.398	**-0.843**	-0.157	-0.262	0.175	0.087
DO	0.205	**0.959**	0.010	0.057	0.111	0.148



(b) Correlation matrix (Spearman rank)

Variables	Flow	Depth	Velocity	Temp	BOD	DO
Flow	**1**	**0.621**	0.423	-0.444	**0.557**	-0.218
Depth	**0.621**	**1**	**0.654**	-0.453	0.425	-0.032
Velocity	0.423	**0.654**	**1**	**-0.760**	0.023	**0.582**
Temp	-0.444	-0.453	**-0.760**	**1**	-0.113	-0.444
BOD	**0.557**	0.425	0.023	-0.113	**1**	**-0.711**
DO	-0.218	-0.032	**0.582**	-0.444	**-0.711**	**1**

Fig. C1. Results of the dataset pre-processing. Evaluation of correlation between predictor variables (i.e. collinearity) with (a) Principal Component Analysis (PCA) and (b) the (Spearman rank (*S*) correlation coefficient.

Appendix C2. Water quality assessment of the Cauca river

The water quality assessment of the Cauca river was performed considering the BMWP-Colombia (Zúñiga and Cardona, 2009), the Dissolved Oxygen Prati (DO-Prati) index (Prati et al., 1971) and an Expert Knowledge Based Index (EKBI) developed by the authors. The EKBI considered different physicochemical water quality classes (Fig. C.2) calculated according to concentration ranges of DO and BOD_5 in rivers reported in literature (NSF, 2003; Ramirez et al., 1999). In order to have a common classification between the ecological and physicochemical indices, the different water quality (WQ) classes and pollution levels were unified according to the following classification: Class 1: unpolluted or very good WQ; Class 2: acceptable pollution level or good WQ; Class 3: moderately polluted or moderate WQ; Class 4: polluted or deficient WQ; Class 5: heavily polluted or bad WQ.

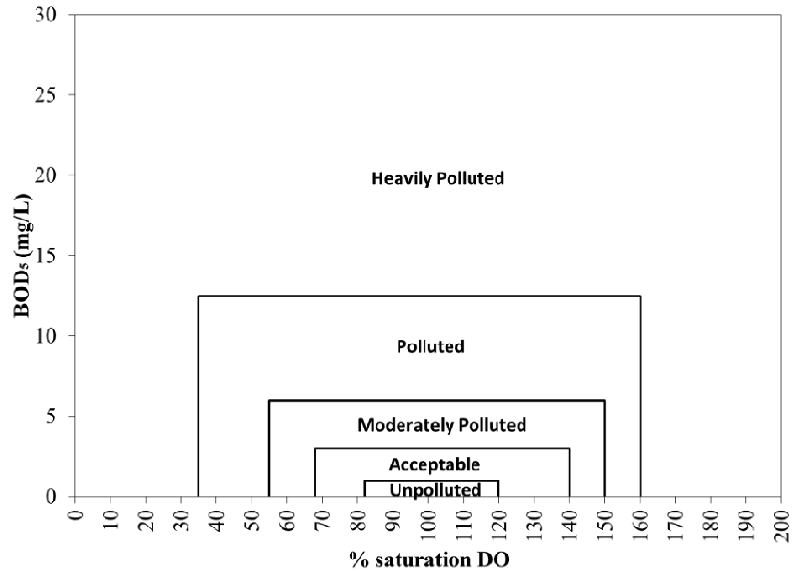


Fig. C.2. Expert knowledge based index (EKBI) developed for the water quality assessment of the Cauca river. The figure is divided in five zones of physicochemical water quality, going from unpolluted to heavily polluted, defined by the concentration of two variables (% of saturation of Dissolved Oxygen (DO) and BOD_5).

The results of the water quality assessment of the Cauca river are presented in Fig. C.3. The spread of the BMWP-Colombia scores over the ecological quality classes of the Cauca river is shown in Fig. C.3a. Additionally, in Figures C.3b - C.3d a comparison between the results of the water quality assessment using ecological and physicochemical indices (EKBI and DO-Prati) is presented. Fig. C.3b shows that the EKBI over-predicts the WQ classes estimated by the ecological index (BMWP-Colombia) in 53 % of the samples, whereas the DO-Prati (Fig. C.3c) over predicts in 73 % of the cases. When the two physicochemical indices are compared (Fig. C.3d), similar results are found in their classifications, thus 53 % of the cases have the same WQ class (delta = 0).

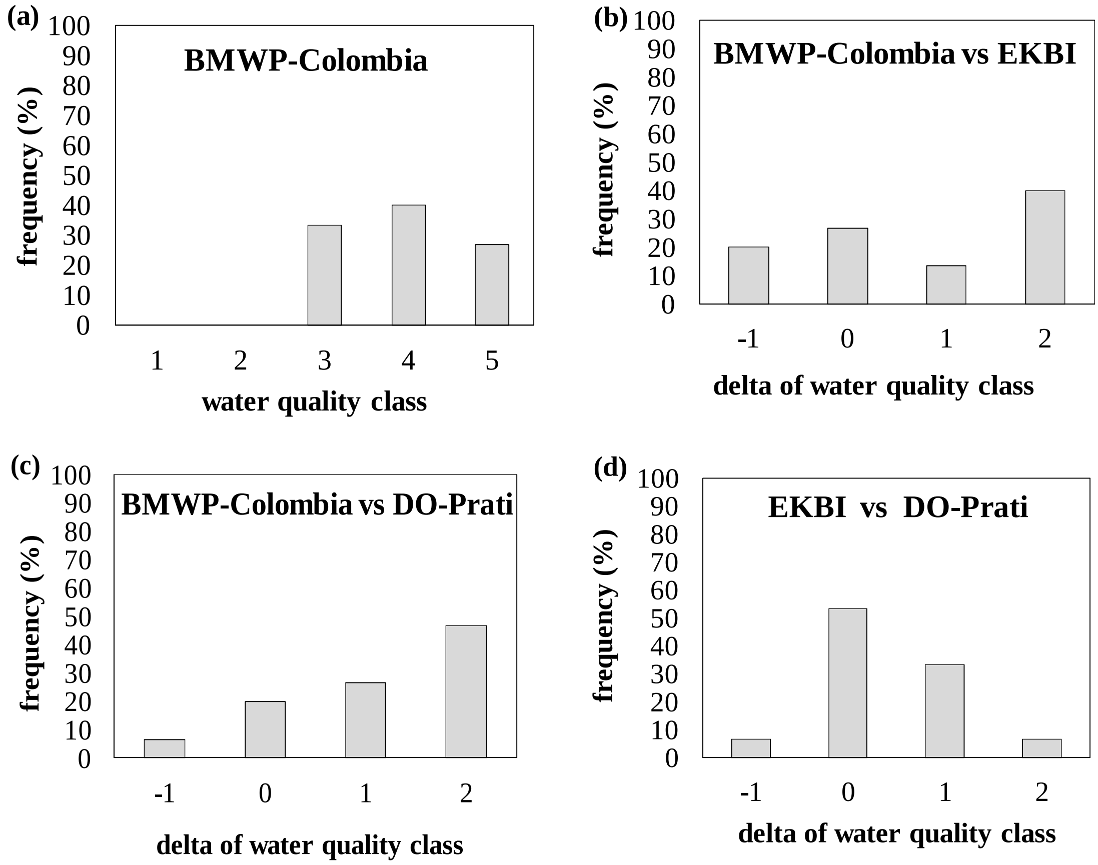


Fig. C.3. Results of the water quality assessment of the Cauca river using ecological and physicochemical indices (a) frequency of Water Quality classes considering the BMWP-Colombia in the dataset; (b) comparison between the BMWP-Colombia and the EKBI; (c) comparison between the BMWP-Colombia and the DO-Prati index; (d) comparison between the EKBI and the DO-Prati index. The abscissa axis (x-axis) of the Fig. C3a corresponds to the EWQ classes 1 to 5 described in the materials and methods section (section 3.2.2) of the Cauca river case study, whereas the abscissa axis of the Fig. C3b – C3d corresponds to the delta of the water classes between the two indices mentioned in the upper part of the graph.

Appendix C3. Multi-model inference based on the information-theoretic approach

The information-theoretic (I-T) approach is based on formulating a series of models that rely on an understanding of the system being studied, followed by an assessment of how the different possible models can be compared to reality (Rushton et al., 2004). Thus, that model (or possibly a small set of models) can be selected as a better approximation of the reality than the rest of the models. The Akaike's information criterion (AIC, Akaike, 1974) forms the basis of I-T approaches in model selection (i.e. identification of variables for inclusion or exclusion in models).

AIC selects a model that fits well but has a minimum number of variables to ensure simplicity and parsimony; consequently, the lower the AIC value the better the model performs (Johnson and Omland, 2004). Additionally, the relative probability of each model being the best model was calculated considering their Akaike weights (w_i). These weights are useful because they: 1) can be used to identify a 95% confidence set of models; 2) provide quantitative information about the support for one model relative to another and; 3) can be used to calculate the relative importance of a variable by summing the w_i of all the models that include that variable (Burnham and Anderson, 2002). All possible combinations of predictor variables were considered to build linear regression models. Interaction of predictors was not considered because the smaller the dataset (n = 15), the more difficult it is to include these terms (Zuur et al., 2009). The maximum number of possible models to evaluate is defined by the number of predictors ($M = 2^P-1$, with P predictors) and it is limited by the sample size (Burnham et al., 2011). Considering the number of predictor variables after the collinearity analysis (three variables, see further) and that sample size should be considerably in excess of the number of predictor variables (Mundry, 2011), seven models were selected to be compared. According to Burnham et al. (2011), AICc differences (Δ_i) lower than four units define models with substantial support for explaining variation in the data. Regarding model performance, if a set of models fit the data poorly, the AICc will only select the best of that poor set. It is therefore necessary to evaluate the models using criteria other than AICc, such as the goodness of fit (Symonds and Moussalli, 2011).

Appendix C4. Diagnostic plots for the model adequacy in the validation of logistic regression model for Ephemeroptera (most parsimonious model, Table 3.3) in the Cauca river

The validation of the GLM models consisted of: (1) a post-hoc evaluation of the model adequacy (Zuur et al., 2009; Fox and Weisberg, 2011) and; (2) the evaluation of the predictive performance of the selected models (Gibson et al., 2004). In this case, it is necessary to evaluate the assumption of the Bernoulli distribution for presence/absence data (the response variable is a vector with ones and zeros) and the Poisson or negative binomial distribution for count data. Plots of residuals versus fitted values and versus each of the predictors were used to detect nonlinear trends, trends in variation across the graph or isolated points (Fox and Weisberg, 2011). A fitted smooth curve helped us to evaluate nonlinear trends. Due to the use of techniques that do not need a normal distribution, a normal distribution of the residuals is no longer of concern (Zuur et al., 2007). Therefore, histograms and QQ-plots of the residuals should be interpreted in terms of how well the model fits the data rather than the normality of the residuals (Zuur et al., 2007). Other diagnostic graphs were used to detect outliers, high-leverage points and influential observations.

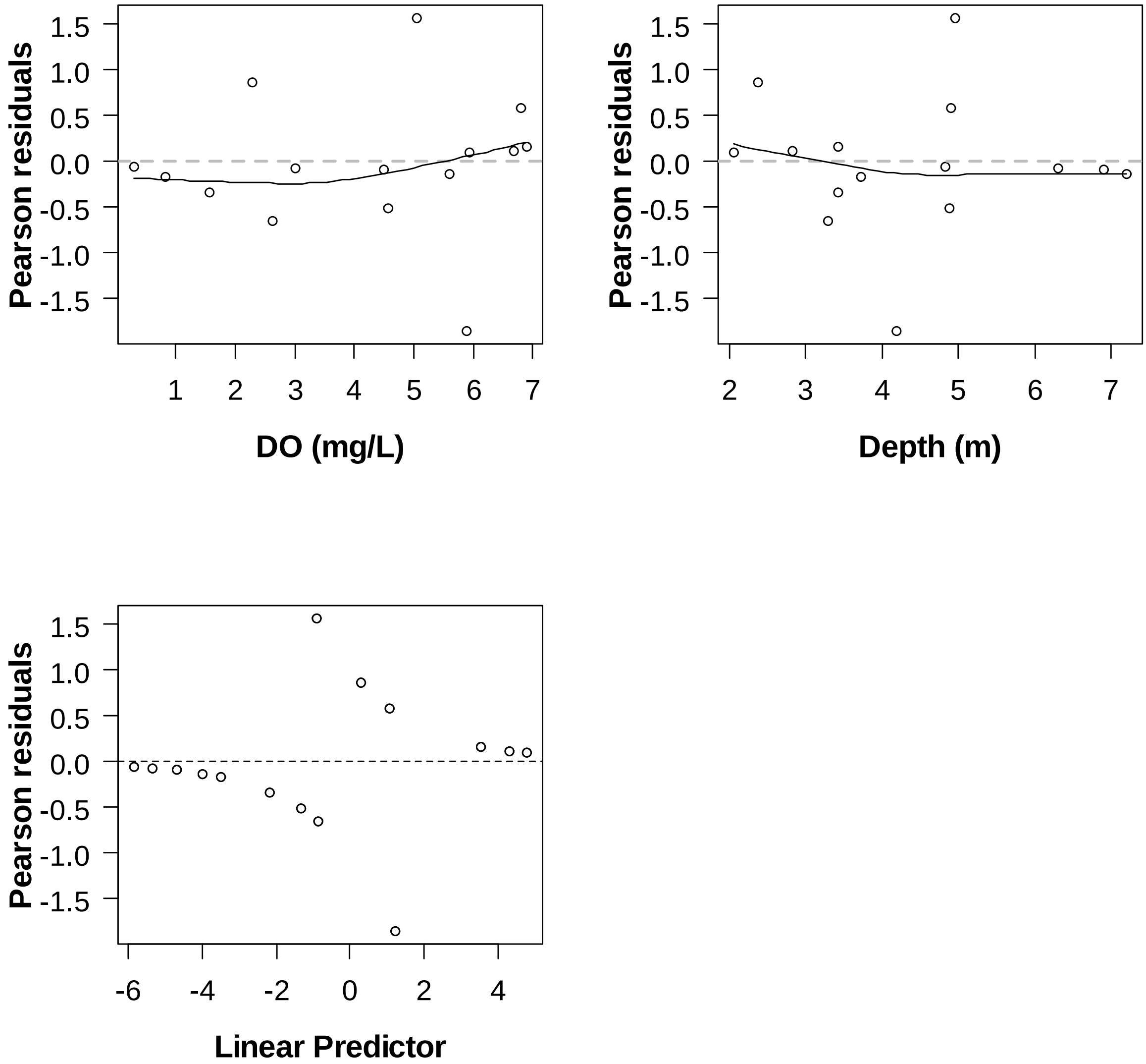


Fig. C.4. Diagnostic plots for the model adequacy in the validation of logistic regression model for Ephemeroptera

Lack-of-fit test available in the package CAR in R (Fox and Weisberg, 2011): in both variables, we have a p-value > 0.05, confirming that these plots do not indicate lack of fit.

Variables	Test statistics	p-value
Depth	0.472	0.492
DO	0.000	0.995

The plot of Pearson residuals against the linear predictor is strongly patterned because the residuals can have only two values, depending on whether the response is equal to zero or one (Fox and Weisberg, 2011).

Appendix C5. Diagnostic plots for the model adequacy in the validation of the logistic regression model for Haplotaxida (most parsimonious model, Table 3.3) in the Cauca river

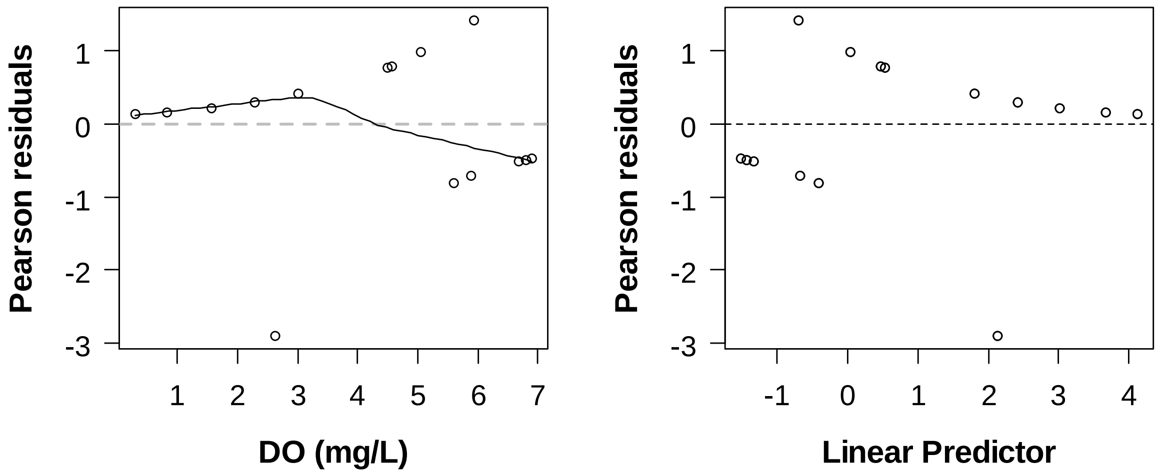


Fig. C.5 Diagnostic plots for the model adequacy in the validation of the logistic regression model for Haplotaxida

Lack-of-fit test available in the package CAR in R (Fox and Weisberg, 2011): we have a p-value>0.05, confirming that this plot does not indicate lack of fit.

Variable	Test statistics	p-value
DO	0.846	0.358

The plot of Pearson residuals against the linear predictor is strongly patterned because the residuals can have only two values, depending on whether the response is equal to zero or one (Fox and Weisberg, 2011).

Appendix C6. Diagnostic plots for the model adequacy in the validation of the negative binomial regression model for the BMWP-Colombia (most parsimonious model, Table 3.3) in the Cauca river

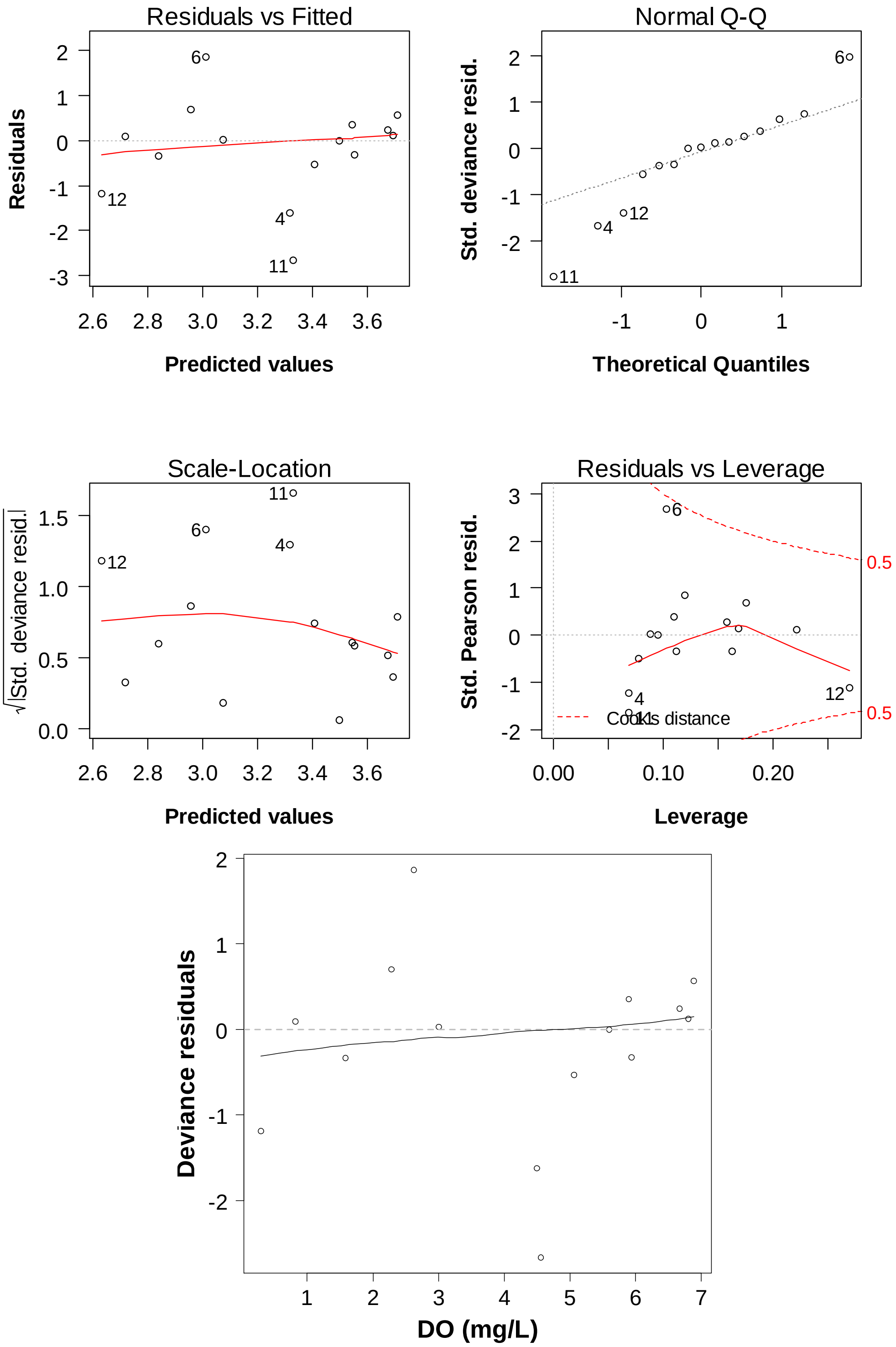


Fig. C.6 Diagnostic plots for the model adequacy with the validation of the negative binomial regression model for the BMWP-Colombia.

Lack-of-fit test available in the package CAR in R (Fox and Weisberg, 2011): we have a p-value > 0.05, confirming that this plot does not indicate lack of fit.

Variable	Test statistics	p-value
DO	-0.014	1.0

Appendix C7. Sensitivity analysis for the logistic regression models and negative regression models implemented in the Cauca river

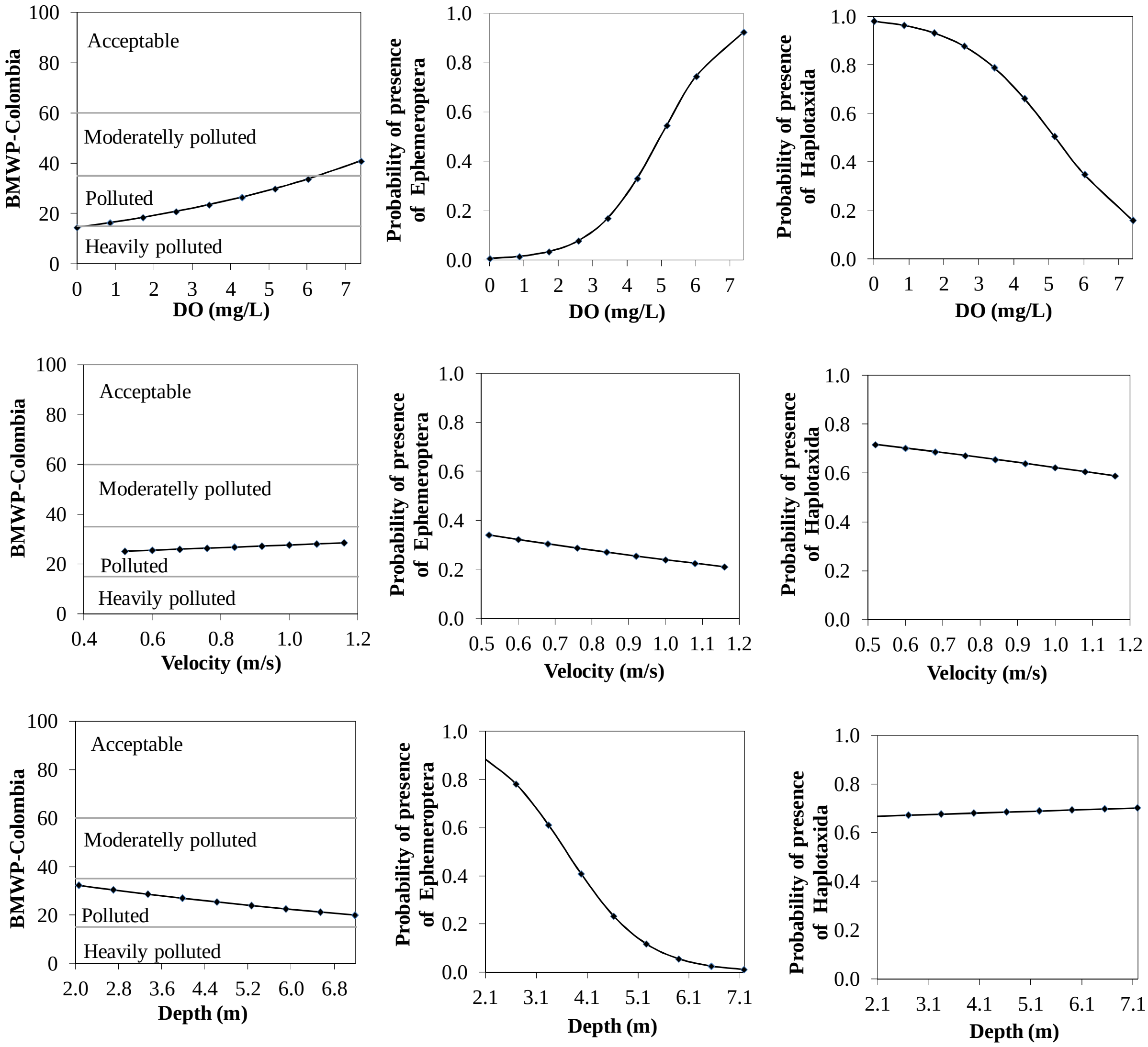


Fig. C.7 Sensitivity analysis for the logistic regression models and negative regression models implemented.

Table C.1 Condition number for de variables included in the regression models during the sensitivity analysis (DO = Dissolved oxygen).

	Condition number		
Model	DO	Velocity	Depth
LRM for Ephemeroptera	2.98	-0.51	-3.49
LRM for Haplotaxida	-0.99	-0.20	0.04
NBRM for BMWP-Colombia	0.58	0.14	-0.40

Appendix D. Supplementary information of the case study of Ecuador (Chapter 4.)

Appendix D1. Dataset pre-processing in the case study of the Cuenca river

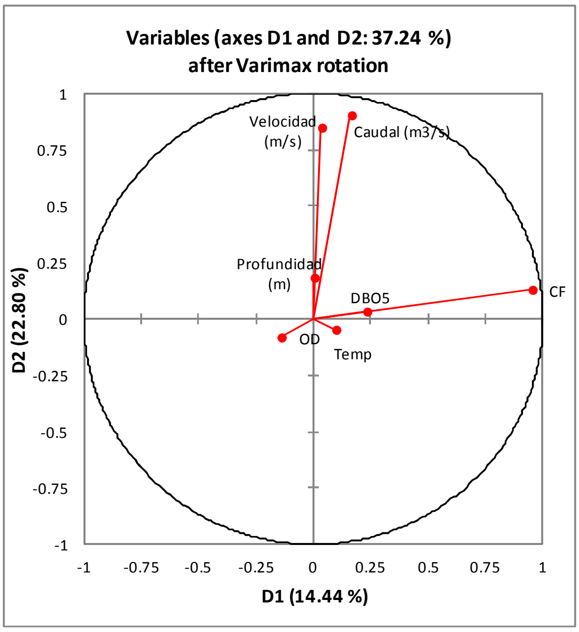
(a) Principal Component Analysis (PCA) with a varimax rotation

Percentage of variance after Varimax rotation:

	D1	D2	D3	D4	D5	D6	F7
Variability (%)	14.44	22.80	14.43	14.62	14.83	14.32	4.57
Cumulative %	14.44	37.24	51.66	66.28	81.11	**95.43**	100.00

Factor loadings after Varimax rotation:

	D1	D2	D3	D4	D5	D6
Temperature	0.094	-0.047	-0.147	**0.954**	-0.051	0.219
DO	-0.144	-0.081	**0.967**	-0.141	-0.089	-0.078
FC	**0.949**	0.129	-0.152	0.095	0.001	0.220
DBO5	0.229	0.033	-0.084	0.230	-0.080	**0.937**
Flow	0.162	**0.903**	-0.030	0.062	0.078	0.108
Velocity	0.032	**0.848**	-0.118	-0.160	0.261	-0.072
Depth	0.000	0.183	-0.086	-0.047	**0.973**	-0.071



(b) Correlation matrix (Kendall)

Variables	Temperature	DO	FC	DBO5	Flow	Velocity	Depth
Temperature	1	-0.294	0.247	0.465	0.025	-0.168	-0.112
DO	-0.294	1	-0.323	-0.213	-0.170	-0.160	-0.176
FC	0.247	-0.323	1	0.464	0.291	0.143	0.015
DBO5	0.465	-0.213	0.464	1	0.162	-0.061	-0.145
Flow	0.025	-0.170	0.291	0.162	1	0.627	0.258
Velocity	-0.168	-0.160	0.143	-0.061	0.627	1	0.404
Depth	-0.112	-0.176	0.015	-0.145	0.258	0.404	1

Fig. D1. Results of the dataset pre-processing. Evaluation of correlation between predictor variables (i.e. collinearity) with (a) Principal Component Analysis (PCA) with a varimax rotation; and (b) the Kendall´s (τ) correlation coefficient.

Appendix E. Supplementary information of the case study of Croatia (Chapter 5.)

Appendix E1. Dataset pre-processing in the case study of the Drava river

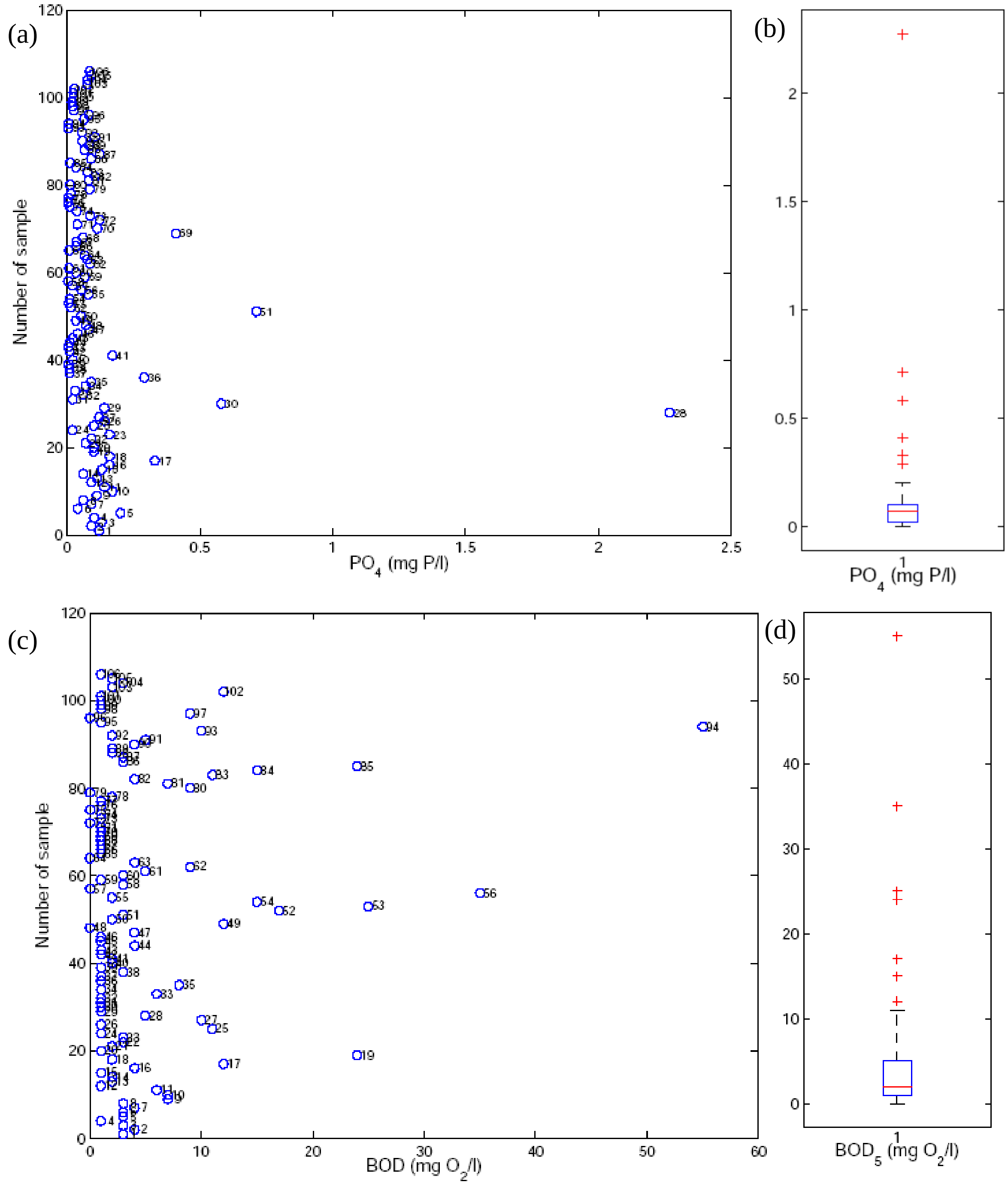


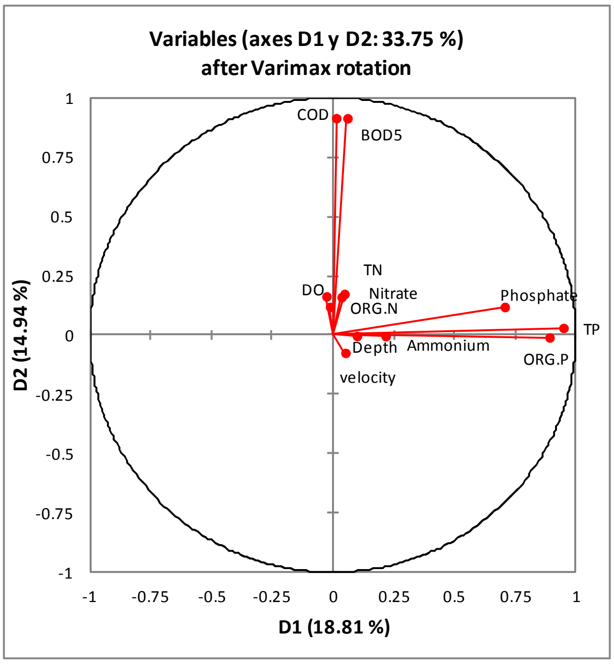
Fig. E1. Results of the dataset pre-processing. Example of the evaluation of outliers with Cleveland dot plots for (a) phosphate-PO_4 and (c) five-day biological oxygen demand-BOD_5 and box plots for (b) PO_4 and (d) BOD_5. See the isolated points at the right side of both Cleveland dot plots that indicates possible outliers.

(a) Percentage of variance after Varimax rotation:

	D1	D2	D3	D4	D5	D6	D7	D8
Variability	18.808	14.940	16.041	9.335	8.656	9.222	8.342	8.680
Cumulative	18.808	33.748	49.789	59.124	67.779	77.002	85.343	94.023

Factor loadings after Varimax rotation:

	D1	D2	D3	D4	D5	D6	D7	D8
DO	-0.028	0.164	-0.100	-0.012	0.191	-0.007	-0.008	0.936
COD	0.013	0.915	0.153	-0.079	0.196	-0.079	0.014	0.017
BOD5	0.059	0.915	0.167	-0.001	0.023	0.080	-0.020	0.186
TN	0.046	0.175	0.946	0.029	0.178	0.097	-0.068	-0.068
Nitrate	0.035	0.161	0.092	-0.107	0.947	0.031	0.028	0.182
TP	0.948	0.031	-0.012	0.127	0.029	0.209	0.089	-0.060
Phosphate	0.706	0.120	-0.139	0.379	0.115	0.271	0.051	-0.264
Ammoniun	0.216	-0.003	-0.056	0.037	0.027	0.963	0.085	-0.006
ORG.N	-0.013	0.120	0.949	0.014	-0.065	-0.177	-0.103	-0.032
ORG.P	0.891	-0.009	0.126	-0.200	-0.033	-0.013	0.034	0.138
Depth	0.098	-0.003	-0.134	0.111	0.025	0.082	0.975	-0.008
velocity	0.051	-0.074	0.055	0.942	-0.118	0.032	0.123	-0.004



(b) Correlation matrix (spearman)

Variables	DO	COD	BOD5	TN	Nitrate	TP	Phosphate	Ammonium	ORG.N	ORG.P	Depth	velocity
DO	**1**	0.197	0.276	-0.086	0.338	-0.059	-0.133	-0.028	-0.088	0.014	0.007	-0.113
COD	0.197	**1**	0.783	0.337	0.342	0.010	0.044	-0.063	0.239	0.061	-0.025	-0.136
BOD5	0.276	0.783	**1**	0.307	0.252	0.087	0.073	0.068	0.245	0.106	-0.032	-0.048
TN	-0.086	0.337	0.307	**1**	0.259	0.066	0.009	0.044	0.885	0.124	-0.171	0.034
Nitrate	0.338	0.342	0.252	0.259	**1**	0.041	0.012	0.064	0.018	0.097	0.026	-0.183
TP	-0.059	0.010	0.087	0.066	0.041	**1**	0.823	0.410	-0.051	0.767	0.217	0.158
Phosphate	-0.133	0.044	0.073	0.009	0.012	0.823	**1**	0.405	-0.118	0.361	0.226	0.273
Ammoniun	-0.028	-0.063	0.068	0.044	0.064	0.410	0.405	**1**	-0.247	0.203	0.190	0.108
ORG.N	-0.088	0.239	0.245	0.885	0.018	-0.051	-0.118	-0.247	**1**	0.050	-0.236	0.008
ORG.P	0.014	0.061	0.106	0.124	0.097	0.767	0.361	0.203	0.050	**1**	0.061	-0.035
Depth	0.007	-0.025	-0.032	-0.171	0.026	0.217	0.226	0.190	-0.236	0.061	**1**	0.208
velocity	-0.113	-0.136	-0.048	0.034	-0.183	0.158	0.273	0.108	0.008	-0.035	0.208	**1**

Fig. E2. Results of the dataset pre-processing. Evaluation of correlation between predictor variables (i.e. collinearity) with (a) Principal Component Analysis (PCA) with a varimax rotation; and (b) the Spearman rank (*S*) correlation coefficient.

Appendix E2. Supplementary material related with the river water quality model implemented in the Drava river

As Chapra (1997) properly pointed out, a continuous stirred tank reactor, is among the simplest systems that can be used to model a water body. Therefore, in this study the conceptual pollutant transport routing was based on the assumption that a river can be represented by a cascade of Continuous Stirred Tank Reactor in Series (CSTRS). The following physicochemical processes were considered during the modelling process: (1) settling processes of organic phosphorus, phosphate and organic matter; (2) hydrolysis of organic nitrogen and organic phosphorus; (3) nitrification and denitrification; (4) decay of organic matter; (5) diffuse pollution and infiltration water processes (infiltrated water from the lake) in the infiltration canal; (6) reaeration. The reaeration rates (k_a) in the Drava river and the canals were calculated as a function of the water depth H (m) and the water velocity U (m/s) as described by Chapra (1997). The k_a values for the Čakovec lake were calibrated in the range between 0 and 2 1/d (Bowie et al., 1985). The following processes were not considered: (1) interactions between sediment layer and water column; (2) algae growth and (3) transport and settling processes of TSS.

Table E1. Processes modelled and calibration ranges of model rate parameters considered (C: Calibration, E: Estimation) (source: Chapra, 1997; Kannel et al., 2007; Cho and Ha, 2010).

Process	Parameter	C/E	Range	
			Min	Max
Settling of organic phosphorus (m/d)	$v_{s,ORGP}$	C	0	2
Settling of phosphate (m/d)	$v_{s,PO4}$	C	0	2
Hydrolysis of organic phosphorus	$k_{d,ORGP}$	C	0.001	0.1
Settling of organic nitrogen(m/d)	$v_{s,ORGN}$	C	0	2
Hydrolysis of organic nitrogen (1/d)	k_{oa}	C	0	5
Nitrification (1/d)	k_{an}	C	0	10
Denitrification (1/d)	k_{dn}	C	0	2
Sink flux for NO_3 (1/d)	$k_{n,s}$	C	0	5
Settling of organic matter (m/d)	$v_{s,ORGC}$	C	0	2
Decay of organic matter (m/d)	$k_{d,ORGC}$	C	0	5
Diffuse organic pollution	L_r	C	0	5
Reareation (1/d)	k_a	C/E	-	-
	DO_{sat}	E	-	-
Nitrogen Oxygen Demand (1/d)	NBOD	E	-	-

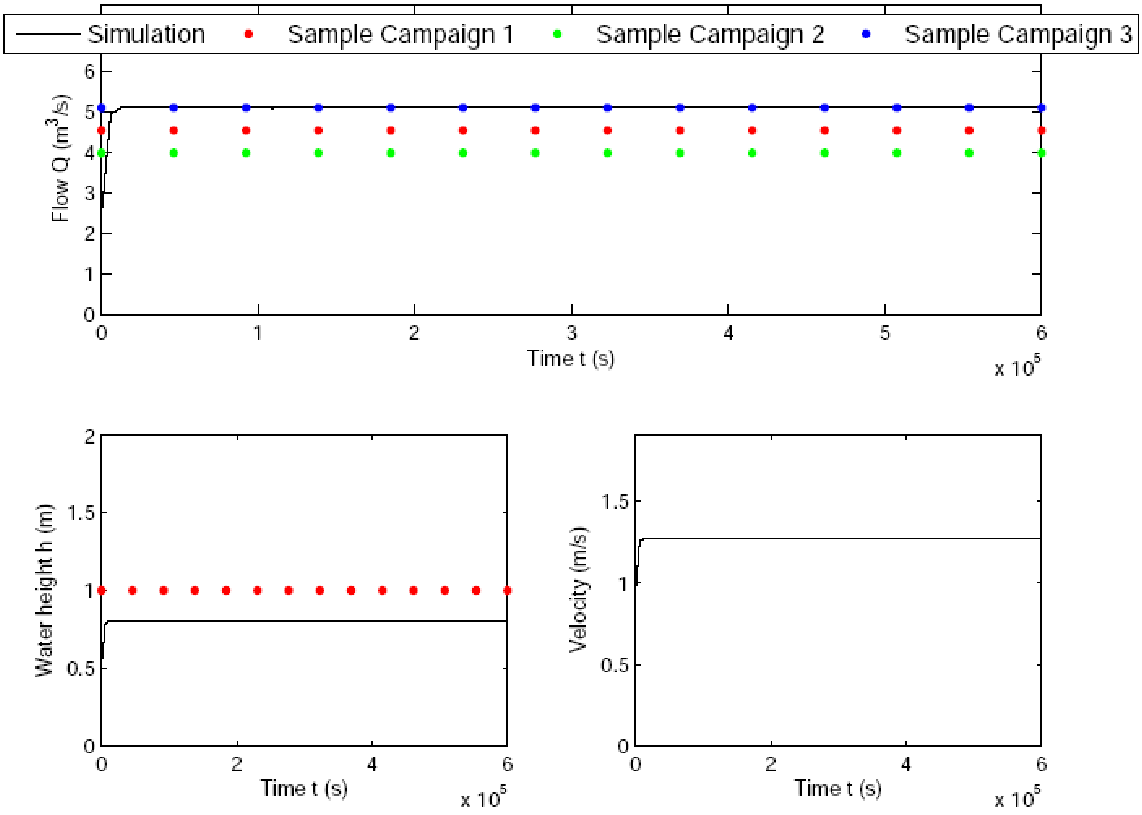


Fig. E3. Example of the calibration for the hydraulic variables in the water quantity model of the Drava river (stretch 5).

Appendix E3. Procedure followed for the construction of the regression trees (RT) in the internal validation procedure of the Drava river

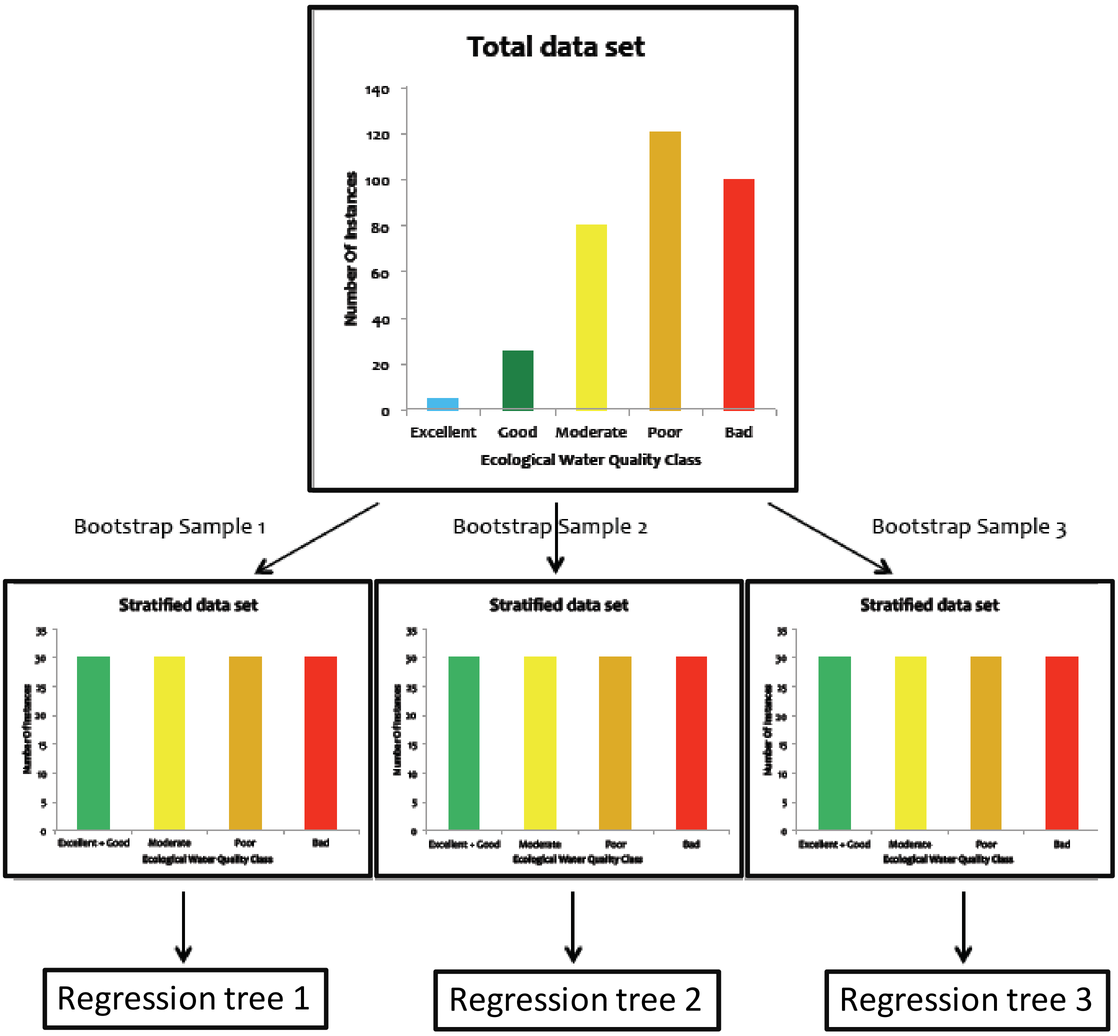


Fig. E4. Example of the bootstrapping technique and the stratification procedure for the dataset in the construction of the regression trees (RT). See that eventually each ecological water quality class, has the same chance to be included in the RT.

References

Abdul-Aziz, O.I., Wilson, B.N., Gulliver, J.S. (2010). Two-zone model for stream and river ecosystems. Hydrobiologia 638, 85–107.

Addinsoft (2010). Statistical software XLSTAT. User Manual. Available at http://www.xlstat.com/en/products/.

Adriaenssens, V. (2004). Knowledge-based macroinvertebrate habitat suitability models for use in ecological river management. PhD thesis, Ghent University, Pags. 296.

Ahmadi-Nedushan, B., St-Hilaire, A., Berube, M., Robichaud, E., Thiemonge, N., Bobee, B. (2006). A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment. River Research and Applications 22, 503-523.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716-723.

Alderink, R.H., Klaver, N.J., Noorman, R. (1995). DUFLOW V 2.0: Micro-computer package for the simulation of 1-dimentional flow and water quality in a network of open water courses. Modelling water quality and flow in river Vecht using DUFLOW. In: Proceedings of the International Conference on Water Quality Modelling, Heatwole C. (ed.), ASAE, Orlando, FL, USA.

Allan J.D. (1995). Stream Ecology: structure and function of running waters. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Allan, J.D., Castillo, M.M. (2007). Stream Ecology. Structure and function of running waters. 2nd ed. Springer, The Netherlands, 436 p.

Araujo M.B., Guisan A. (2006). Five (or so) challenges for species distribution modelling. Journal of Biogeography 33 (10), 1677–1688

Aspinall, R. (2002). Use of logistic regression for validation of maps of the spatial distribution of vegetation species derived from high spatial resolution hyperspectral remotely sensed data. Ecological Modelling 157, 301-312.

Aspinall, R. (2004). Modelling land use change with generalized linear models--a multi-model analysis of change between 1860 and 2000 in Gallatin Valley, Montana. J. Environmental Management 72, 91-103.

Bauwens, W. (2009). Course notes Water Quality Management. Academic Year 2008-2009. Faculty of Bioscience Engineering, Ghent University, Belgium.

Beck, M.B., Reda, A. (1994). Identification and application of a dynamic model for operational management of water quality. Water Science and Technology 30 (2), 31–41.

Beecher, H.A., Caldwell, B.A., Demond, S.B. (2002). Evaluation of depth and velocity preferences of juvenile coho salmon in Washington streams. North American Journal of Fisheries Management 22: 785–795.

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

Beven, K.J., Binley, A. (1992). The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes 6, 279-298.

Bockelmann, B.N., Fenrich, E.K., Lin, B., Falconer, R.A. (2004). Development of an eco-hydraulics model for stream and river restoration. Ecological Engineering 22, 227-235.

Boets, P., Lock, K., Goethals, P.L.M. (2010). Combining data-driven methods and lab studies to analyse the ecology of *Dikerogammarus villosus*. Ecological Informatics 5, 133-139.

Boets P., Holguin-Gonzalez J.E., Lock, K., Goethals, P.L.M. (2013). Data-driven habitat analysis of the Ponto-Caspian amphipod Dikerogammarus villosus in two invaded regions in Europe. Ecological Informatics 17, 36-45. doi:10.1016/j.ecoinf.2012.07.001

Boets, P., Lock K., Goethals P.L.M. (in press a). Modelling habitat preference, abundance and species richness of alien macrocrustaceans in surface waters in Flanders (Belgium) using decision trees. Ecological Informatics, doi: 10.1016/j.ecoinf.2012.06.001.

Boets, P., Pauwels, I.S., Lock, K., Goethals, P.L.M. (in press b). Using an integrated modelling approach for risk assessment of the 'killer shrimp' *Dikerogammarus villosus*. River Research and Applications, doi: 10.1002/rra.2658

Borchart, D., Sperling, F. (1997). Urban stormwater discharges: Ecological effects on receiving waters and consequences for technical measures, Water Science and Technology 36 (8-9), 173-178.

Bovee, K.D., Lamb, B.L., Bartholow J.M., Stalnaker, C.B., Taylor, J., Henriksen, J. (1998). Stream habitat analysis using the Instream Flow Incremental Methodology. U.S. Geological Survey, Biological Resources Discipline Information and Technology Report USGS/BRD-1998-0004.

Bowie, G.L, et al. (1985). Rates, constants, and kinetics formulations in surface water quality modeling. Rep. No. EPA/600/3-85/040, second edition, U.S. Environmental Protection Agency, Georgia, U.S.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1984). Classification and regression trees. Wadsworth International: Belmont, p. 358.

Breur, K.-J, van Leeuwen, P.E.R., Dellaert, N.P. (1997). Deriving discharge strategies to reduce CSO in urban drainage systems, Water Science and Technology 36 (8-9), 259-263.

Brink B., et al .(1991). AMOEBA, Approach as a Useful Tool for establishing Sustainable Development. Final report. pp 71-88.

Brown, L., Barnwell, T. (1987). The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS: Documentation and user manual. United States Environmental Protection Agency -EPA. Georgia. U.S.

Burnham, K.P., Anderson, D.R. (2002). Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach, 2nd edn. Springer, New York, United States.

Burnham, K.P, Anderson, D.R, Huyvaert, K.P. (2011). AICc model selection in Ecological and behavioral science: some background, observations, and comparisons. Behavioral Ecology and Sociobiology 65, 23-35

Bwalya, C.Y. (1996). Model based evaluation of operating strategies in the sewer-waste water treatment plant-river system, M.Sc. Thesis, Faculty of agricultural and applied biological sciences, University of Gent, Belgium.

Camacho, L.A., Lees, M.J. (1999). Multilinear discrete lag-cascade model for channel routing. Journal of Hydrology 226, 30–47.

Camacho, L.A., González, R.A. (2008). Calibration and predictive ability analysis of longitudinal solute transport models in mountain streams. Environmental Fluid Mechanics 8, 597–604.

Carpenter, D.A. (2001). Testing the river continuum concept: exploring ecological processes in the wheeling creek watershed. Master thesis, Wheeling Jesuit University, Morgantown, West Virginia.

Carrasco, M.C. (2008). Biotic integrity and its application, a proposal for the management of the rivers Tomebamba and Yanuncay in Cuenca, Ecuador. Master thesis, Faculty of Natural sciences. Universidad Autónoma de Querétaro, México.

Cerco, C.F., Cole, T. (1995). User guide to the CE-QUAL-ICM three dimentional eutrophication model. Release version 1.0, Technical reports EL-95-15, US Army Eng., Waterways Experiment Station, Vicksburg, MS.

Chapman, Deborah (1996). Water quality assessments. A Guide to the use of biota, sediments and water in environmental monitoring, Second edition. Published on behalf of UNESCO, WHO and UNEP, Chapman & Hall, London, UK, 651 pp.

Chapra S.C., Pelletier G.J., Tao H. (2008). QUAL2K: A Modeling Framework for Simulating River and Stream Water Quality, Version 2.11: Documentation and Users Manual. Civil and Environmental Engineering Dept., Tufts University, Medford, MA.,

Chapra, S. C. (1997). Surface Water-quality Modeling. McGraw-Hill, New York, U.S.

Chapra, S.C., Pelletier G.J. (2003). QUAL2K: A Modeling Framework for Simulating River and Stream Water Quality (Beta Version): Documentation and Users Manual. Civil and Environmental Engineering Dept., Tufts University, Medford, MA.

Cho, J.H., Ha, S.R. (2010). Parameter optimization of the QUAL2K model for a multiple-reach river using an influence coefficient algorithm. Science of the Total Environment 408, 1985-1991.

Chow, V.T. (1981). Open-channel Hydraulics. McGraw Hill Book Company, Inc, New York.

Chow, V.T., Maidment, D.R., Mays, L.W. (1988). Applied Hydrology. McGraw Hill Book Company, Inc, New York.

Churchill, M.A., Elmore, H.L., Buckingham, R.A. (1962). The prediction of stream reaeration rates. Journal of the Sanitary Engineering Division, ASCE, 88, 1-46.

Clark, M.J. (2002). Dealing with uncertainty: adaptive approaches to sustainable river management. Aquatic Conservation: Marine and Freshwater Ecosystems 12, 347–363.

Cluckie I.D., Xuan Y. (2008). Uncertainty Propagation in Ensemble Rainfall Prediction Systems used for Operational Real-Time Flood Forecasting. Chapter 31 In: Abrahart, R.J., See, L.M., Solomatine, D.P. (Eds.). Practical Hydroinformatics. Computational Intelligence and Technological Developments in Water Applications. Springer. p. 437-447.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20, 37–46.

Cook, S. (1976). Quest for an index of community structure sensitive to water pollution. Environmental Pollution 11(4), 269-288.

CVC- Corporación Autónoma Regional del Valle del Cauca, Univalle-Universidad del Valle. (2007). Cauca River Modelling Project (CRMP), compilation Volumes IX and XII, Phase II, and Volumes XII and XIII, Phase III, (in Spanish). Cali, Colombia.

De Pauw, N., Gabriels, W., Goethals, P.L.M. (2006). River monitoring and assessment methods based on macroinvertebrates, in: Ziglio G, Siligardi M., Flaim G. (Eds.), Biological Monitoring of Rivers. Water Quality Measurements Series. John Wiley & Sons, Ltd., England, pp. 113-133.

De Pauw, N., Hawkes, H.A. (1993). Biological monitoring of river water quality. In: Walley, W.J. & Judd, S. (eds.). River water quality monitoring and control. Aston University, Birmingham, UK. p. 87-111.

De Pauw, N., Vanhooren, G. (1983). Method for biological quality assessment of watercourses in Belgium. Hydrobiologia, 100, 153-168.

De St. Venant, M. (1871). Theory of unsteady water flow, with application to river floods and to propagation of tides in river channels. French Academy of Science, 73, 148–154, 37–240.

De'ath, Glenn (2002). Multivariate regression trees: a new technique for modelling species environment relationships. Ecology 83, 1105-1117.

Debeljak, M., Džeroski, S. (2011). Decision trees in ecological modelling. In: Jopp, F., Reuter, H., Breckling, B., (eds), Modelling complex ecological dynamics, Springer Berlin Heidelberg: 197-209.

Deksissa T. (2004). Dynamic integrated modelling of basic water quality and fate and effect of organic contaminants in rivers. Doctoral Thesis in Applied Biological Sciences. Ghent University. Gent, Belgium.

Deksissa, T., Meirlaen, J., Ashton, P., Vanrolleghem, P. (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.

Deltares (2009). Water Framework Directive Explorer (WFD-Explorer) manual. The Netherlands.

Devesa, F., Comas, J., Turon, C., Freixó, A., Carrasco, F., Poch, M. (2009). Scenario analysis for the role of sanitation infrastructures in integrated urban wastewater management. Environmental Modelling & Software, 24 (3), 371-380.

DHI - Danish Hidraulic Institute (1999). MIKE 11 a modelling system for rivers and channel, short description, guide to getting started, tutorial. DHI—Water and Development, Horsholm, Denmark.

Díaz-Granados, M., Rodríguez, J.P., Rodríguez, M.S., Penagos, J.C., Camacho, L.A., Achleitner, S., Maksimovic, C., McIntyre, N. (2009) Towards a paradigm shift in urban drainage management and modelling in developing countries, Revista de Ingeniería Facultad de Ingeniería Universidad de los Andes, 30, pp. 133-150 ISSN 0121-4993.

Dominguez, G.L., Lock, K., Goethals, P.L.M. (2011a). Application of classification trees to determine biological and chemical indicators for river assessment: case study in the Chaguana watershed (Ecuador). Journal of Hydroinformatics 13, 489-499.

Dominguez, G.L., Lock, K., Goethals, P.L.M. (2011b). 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.

EEA-European Environment Agency (1999). Environmental Indicators: Typology and Overview. European Environment Agency, Technical Report No. 25, Copenhagen, p 19.

Efron, B., Tibshirani, R. (1993). An introduction to the Bootstrap. Chapman and Hall, New York.

ETAPA - Water supply and sanitation company in the city of Cuenca (2007). Water quality evolution in the lowland stretches of the rivers of city of Cuenca 2007 (in Spanish). Environmental management direction.

ETAPA - Water supply and sanitation company in the city of Cuenca (2009). Assessment of the wastewater treatment plant - Period November 1999 - December 2009 (in Spanish). Environmental management direction.

European Commission (2000). Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a Framework for Community action in the field of water policy. Official Journal of the European Communities L327: 1-72

Everaert, G., Boets, P., Lock, K., Goethals, P.L.M. (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., Pauwels, I.S., Boets, P., Buysschaert, F., Goethals, P.L.M. (2012). Development and assessment of ecological models in the context of the European Water Framework Directive: key issues for trainers in data-driven modelling approaches. Ecol. Inform., http://dx.doi.org/10.1016/j.ecoinf.2012.10.007.

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

Extence, C., Balbi, D., Chadd, R. (1999). River flow indexing using British benthic macroinvertebrates: A framework for setting hydroecological objectives. Regulated rivers-research and management, 15(6): 543-574.

Fielding, A.H. (2002). What are the appropriate characteristics of an accuracy measure? In: Scott, J.M., Heglund, P., Morrison, M., Raven, P., (Eds.), Predicting species occurrences: issues of accuracy and scale. Island Press, Covelo, CA., pp. 271-280.

Fielding, A.H., Bell, J.F. (1997). A review of methods for the assessment of prediction errors in conservation presence–absence models. Environ. Conserv. 24, 38-49.

Fitzpatrick, M., Weltzin, J., Sanders, N., Dunn, R. (2007). The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? Global Ecology and Biogeography 16, 24–33.

Fox, J., Weisberg S. (2011). An R companion to applied regression, Second Edition. Sage Publications, California, U.S.

Freckleton, R.P. (2011). Dealing with collinearity in behavioural and ecological data: model averaging and the problems of measurement error. Behavioral Ecology and Sociobiology 65, 91–101.

Gabriels, W. (2007). Multimetric assessment of freshwater macroinvertebrate communities in Flanders, Belgium. PhD thesis. Faculty of Bioscience Engineering, Ghent University, Belgium.

Gabriels, W., Goethals, P.L.M., Dedecker, A., Lek, S., De Pauw, N. (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., Lock K., De Pauw, N., Goethals P.L.M. (2010). Multimetric Macroinvertebrate Index Flanders (MMIF) for biological assessment of rivers and lakes in Flanders (Belgium), Journal of Limnology 40, 199–207.

Gal, G., Skerjanec, M., Atanasova, N. (2013). Fluctuations in water level and the dynamics of zooplankton: a data-driven modelling approach. Freshwater Biology 58, 800-816.

Gernaey, K.V., Jørgensen, S.B. (2004). Benchmarking combined biological phosphorus and nitrogen removal wastewater treatment processes. Control Engineering Practice 12 (3), 357-373.

Gibson, L.A., Wilson, B.A., Cahill, D.M., Hill, J. (2004). Spatial prediction of rufous bristlebird habitat in a coastal heathland: a GIS-based approach. Journal of Applied Ecology 41, 213–223.

Goethals P.L.M. (2005). Data driven development of predictive ecological models for benthic macroinvertebrates in rivers. Doctoral Thesis in Applied Biological Sciences. Ghent University. Gent, Belgium.

Goethals, P.L.M. and De Pauw, N. (2001). Development of a concept for integrated ecological river assessment in Flanders, Belgium. Journal of Limnology, 60 (1), 7-16.

Gordon N.D., MacMahon T.A., Finlayson B.L., Gippel C.J., Nathan R.J. (2004). Stream Hydrology: An Introduction for Ecologists, 2nd edn. Wiley, London.

Gore James, Layzer James B. et al. (2001). Macroinvertebrate instream flow studies after 20 years: a role in stream management and restoration." Regulated Rivers: Research & Management 17(4-5): 527-542.

Grian, L., Kerea, Z. (2004). Restoration of the old Drava riverbed at the Čakovec hydropower plant. In 3rd European Conference on River Restoration, Zagreb, Croatia.

Grueberg, C.E., Nakagawa, S., Laws, R.J., Jamieson, I.G. (2011). Multimodel inference in ecology and evolution: challenges and solutions. Journal of Evolutionary Biology 24, 699–711.

Guisan, A., Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. Ecological Modelling, 135, 147-168.

Guisan A. (2002). Semiquantitative response models for predicting the spatial distribution of plant species, in: Scott, J.M., Heglund, P., Morrison, M., Raven, P., (Eds.), Predicting species occurrences: issues of accuracy and scale. Island Press, Covelo, CA., pp. 315-326.

Guisan A., Graham C.H., Elith J., Huettmann F., NCEAS Species Distribution Modelling Group (2007). Sensitivity of predictive species distribution models to change in grain size. Diversity and Distributions 13, 332–340.

Guisan Antoine, Edwards Thomas C. Jr., Hastie Trevor (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157, 89- 100.

Harmel, R.D., Smith, P.K. (2007). Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modelling. Journal of Hydrology 337, 326-336.

Henze M., Gujer W., Mino T., van Loosdrecht M. (2000). Activated sludge models ASM1, ASM2, ASM2D and ASM3. Scientific and technical report No. 9, IWA Publishing, London, UK. Vol 9 (1), p. 121.

Hosmer, D., Lemeshow, S. (2000). Applied Logistic Regression. Second Edition. A. Wiley-Interscience Publication, John Wiley and Sons Inc, New York, NY. 373 pp.

Hughes, D.A., Louw, D. (2010). Integrating hydrology, hydraulics and ecological response into a flexible approach to the determination of environmental water requirements for rivers. Environmental Modelling & Software 25, 910-918.

Hurvich, C.M., Tsai, C.L. (1989). Regression and time series model selection in small samples. Biometrika 76, 297–307.

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

Hynes, H.B.N. (1960). The Biology of Polluted Waters. Liverpool Univ. Press, Liverpool, England. 202 p.

Irvine K., Mills P., Bruen M., Walley W., Hartnett M., Black A., Tynan S., Duck R., Bragg O., Rowen J., Wilson J., Johnston P., O'Toole C. (2002). Environmental RTDI Programme 2000–2006. Water Framework Directive–The Application of Mathematical Models as Decision-Support Tools. An Assessment of Mathematical Modelling in its Implementation in Ireland (2002-W-DS-11). Dublin. Ireland.

Jähnig, S.C., Kuemmerlen, M., Kiesel, J., Domisch, S., Cai, Q., Schmalz, B., Fohrer N., (2012). Modelling of riverine ecosystems by integrating models: conceptual approach, a case study and research agenda, Journal of Biogeography, 39 (12), 2253–2263.

Johnson, J. B., Omland K.S. (2004). Model selection in ecology and evolution. Trends in Ecology & Evolution 19, 101-108.

Jorgensen S.E., Bendoricchio G. (2001). Fundamentals of Ecological Modelling, 3rd Edition. Elsevier. The Netherlands. 530 pages.

Jorgensen S.E., Fath B.D. (2011). Fundamentals of Ecological Modelling. Applications in Environmental Management and Research. 4th edition. Elsevier. The Netherlands. 399 pages.

Jowette, I.G. (2003). Hydraulic constraints on habitat suitability for benthic invertebrates in gravel-bed river. River Research and Applications 19, 495–507.

Kaeiro, K., Moels, T., Timm, H., Virro, T., Jaervekuelg, R. (2011). The effect of damming on biological quality according to macroinvertebrates in some Estonian streams, Central-Baltic Europe: a pilot study. River Research and Applications, 27(7):895-907.

Kannel, P.R., Lee, S., Lee, Y-S., Kanel, S.R., Pelletier, G.J. (2007). Application of automated QUAL2Kw for water quality modeling and management in the Bagmati River, Nepal. Ecological Modelling 202, 503-517.

Kemp, J., Harper, D., Crosa, G. (2000). The habitat-scale ecohydraulics of rivers. Ecological Engineering, 16 (1):17-29.

Kocev, D., Džeroski, S., White, M.D., Newell, G.R, Griffoens, P. (2009). Using single- and multi-target regression trees and ensembles to model a compound index of vegetation condition. Ecological Modelling 220, 1159-1168.

Kraft, Dietmar (2011). Model Integration: Application in Ecology and for Management. In: Jopp, F., Reuter, H., Breckling, B., (eds), Modelling complex ecological dynamics, Springer Berlin Heidelberg, 301-320.

Krause, P., Boyle, D.P, Base, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. 5, 89-97.

Lam, D., Leon, L., Hamilton, S., Crookshank N., Bonin D., Swayne D. (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.

Landuyt, D., Broekx, S., D'hondt R., Engelen G., Aertsens, J., Goethals P.L.M. (in press). A review of Bayesian belief networks in ecosystem service modelling. Environmental Modelling & Software, http://dx.doi.org/10.1016/j.envsoft.2013.03.011

Larocque, G.R., Bhatti, J.S., Ascough, J.C., Liu, J., Luckai, N., Mailly, D., Archambault, L., Gordon, A.M. (2011). An analytical framework to assist decision makers in the use of forest ecosystem model predictions. Environmental Modelling & Software 26, 280–288.

Laws, E.A. (2000). Aquatic pollution: an introductory text. John Wiley & Sons.

Leclerc, M., St-Hilaire, A., Bechara, J.A. (2003). State-of-the-Art and Perspectives of Habitat Modelling for Determining Conservation Flows. Canadian Water Resources Journal 28 (2), 153–172.

Legates, D.R., McCabe, Jr., G.J. (1999). Evaluating the use of ''goodness-of-fit'' measures in hydrologic and hydroclimatic model validation. Water Research 35, 233-241.

Lek, Sovan (2007). Uncertainty in ecological models. Editorial. Ecological Modelling 207, 1-2.

Liu, Y.Q., Gupta, H., Springer, E., Wagener, T., 2008. Linking science with environmental decision making: Experiences from an integrated modeling approach to supporting sustainable water resources management. Environ. Model. Softw. 23 (7), 846-858.

Lock, K., Asenova, M., Goethals, P.L.M. (2011). Benthic macroinvertebrates as indicators of the water quality in Bulgaria: A case-study in the Iskar river basin. Limnologica 41, 334– 338.

Lock, K., Goethals, P.L.M. (2011). Distribution and ecology of the mayflies (Ephemeroptera) of Flanders (Belgium). Annales de Limnologie - International Journal of Limnology 47, 59–165.

Lock, K., Goethals, P.L.M. (in press). Habitat suitability modelling for mayflies (Ephemeroptera) in Flanders (Belgium). Ecological Informatics, doi: 10.1016/j.ecoinf.2011.12.004.

Lorenz C. M., Van Dijk G. M., et al. (1997). Concepts in river ecology: implications for indicator development. Regulated Rivers: Research & Management 13(6): 501-516.

Maddock, I. (1999). The importance of physical habitat assessment for evaluating river health. Freshwater Biology 41(2), 373-391.

MAFRD-Ministry of Agriculture, Fisheries and Rural Development (2009). (in Croatian) Arkod @online.url. http://preglednik.arkod.hr/ARKOD-Web/.

Maggini, R., Lehmann, A., Zimmermann, N.E., Guisan, A. (2006). Improving generalized regression analysis for the spatial prediction of forest communities. Journal of Biogeography 33, 1729–1749.

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

MathWorks (2010). MATLAB technical documentation, http://www.mathworks.com/, Retrieved 2010-06-07.

McCullogh, P., Nelder, J.A. (1989). Generalized Linear Models, 2nd ed. Chapman and Hall, London.

McPherson, J.M., Jetz, W. (2007). Effects of species' ecology on the accuracy of distribution models. Ecography 30, 135–151.

Meirlaen, J., Huyghebaert, B., Sforzi, F., Benedetti, L., Vanrolleghem, P.A. (2001). Simultaneous simulation of the integrated urban wastewater system using mechanistic surrogate models. Water Science and Technology 43 (7), 301-309.

Metcalfe, J. (1989). Biological water quality assessment of running waters based on macroinvertebrate communities. History and present status in Europe. Environmental Pollution, 60, 101-139.

Meynard, C.N., Quinn, J.F. (2007). Predicting species distributions: a critical comparison of the most common statistical models using artificial species. Journal of Biogeography 34, 1455–1469.

Molle, F. (2009). River-basin planning and management: The social life of a concept, Geoforum, 40, 484-494.

Møller, A.P., Jennions, M.D. (2002). How much variance can be explained by ecologists and evolutionary biologists? Oecologia 132, 492–500.

Mouton, A.M., Van Der Most, H., Jeuken, A., Goethals, P.L.M., De Pauw, N. (2009a). 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.

Mouton, A.M., De Baets, B., Van Broekhoven, E., Goethals, P.L.M. (2009b). Prevalence adjusted optimisation of fuzzy models for species distribution. Ecological Modelling 220, 1776-1786.

Mouton, A.M., Jowett, I., Goethals, P.L.M., De Baets, B. (2009c). Prevalence-adjusted optimisation of fuzzy habitat suitability models for aquatic invertebrate and fish species in New Zealand. Ecological Informatics 4, 215–225.

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

Mouton, A.M., De Baets, B., Goethals, P.L.M. (2010). Ecological relevance of performance criteria for species distribution models. Ecological Modelling 221, 1995-2002.

Mullis, R., Revitt, D.M., Shutes R.B.E. (1997). The impacts of discharges from two combined sewer overflows on the water quality of an urban watercourse, Water Science and Technology 36 (8-9), 195-199.

Mundry, R. (2011). Issues in information theory based statistical inference - a commentary from a frequentist's perspective. Behavioral Ecology and Sociobiology 65, 57-68.

Munn, M.D., Brusven, M.A. (1991). Benthic macroinvertebrate communities in nonregulated and regulated waters of the clearwater river. Regul. Rivers Res. Manage. 6, 1–11.

Newson, M., Sear, D., Soulsby, C. (2012). Incorporating hydromorphology in strategic approaches to managing flows for salmonids. Fisheries Management and Ecology, pp. 1365-2400.

NSF-National Sanitation Foundation (2003). Water Quality Index. Boulder Area Sustainability Information Network http://pathfinderscience.net/stream/cproto4.cfm

O´Connor, D.J., Dobbins, W.E., 1956. Mechanism of reaeration in natural streams. ASCE Trans 86, 35-55.

Oddi A., Policella N., Cesta A., Cortellessa G. (2005). Constraint-Based Random Search for Solving Spacecraft Downlink Scheduling Problems. In: Kendall G., Burke E., Petrovic S., Gendreau M. (Eds.), Multidisciplinary Scheduling: Theory and Applications. pp 133-160

O'Kane J.P. (2008). Hydroinformatics – The Challenge for Curriculum and Research, and the "Social Calibration" of Models, Chapter 33. In: Abrahart, R.J., See, L.M., Solomatine, D.P. (Eds.). Practical Hydroinformatics. Computational Intelligence and Technological Developments in Water Applications. Springer. p. 465-478.

Ongley E. D. and Booty W.G. (1999). Pollution remediation planning in developing countries: Conventional modelling versus knowledge based prediction. Water International 24, 31-38.

Owens, M., Edwards, R.W., Gibbs, J.W. (1964). Some reaeration studies in streams. International Journal of Air and Water Pollution 8, 469–486.

Palmer, M.A, Reidy L.C., Nilsson, C., Flörke M., Alcamo, J., Lake, P.S., Bond, N. (2008). Climate change and the world's river basins: anticipating management options. Frontiers in Ecology and the Environment 6, 81–89. doi:http://dx.doi.org/10.1890/060148

Paredes-Arquiola, J., Martinez-Capel F., Solera A., Aguilella V. (2011). Implementing environmental flows in complex water resources systems – case study: the Duero river basin, Spain. River Research and Applications doi: 10.1002/rra.1617.

Pauwels, I.S., Everaert, G., Goethals, P.L.M. (2010). Integrated river assessment by coupling water quality and ecological assessment models. In: Swayne, D.A., Yang, W., Voinov, A.A., Rizzoli, A., Filatova, T. (Eds.), 5th Biennial meeting of the Intern. Congr. on Environ. Model. and Softw. (iEMSs 2010): Modelling for environment's sake, Intern. Environ. Model. and Softw. Society (iEMSs) Ottawa, ON, Canada. http://www.iemss.org/iemss2010/Volume1.pdf, 876-884.

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

Pearson, R.G., Thuiller, W., Araujo M.B., Brotons, L., Martinez-Meyer, E., McClean, C., Miles, L, Segurado, P., Dawson, T.P., Lees, D. (2006). Model-based uncertainty in species range prediction. Journal of Biogeography 33, 1704–1711.

Pelletier, G.J., Chapra, C.S., Tao, H. (2006). QUAL2Kw, A framework for modeling water quality in streams and rivers using a genetic algorithm for calibration. Environmental Modelling & Software 21, 419-4125.

Pelletier, G.J., Chapra, S.C. (2005). QUAL2Kw theory and documentation (version 5.1), A Modeling Framework for Simulating River and Stream Water Quality, retrieved 15 October 2009 from: http://www.ecy.wa.gov/programs/eap/models/.

Pesch, R., Schröder, W. (2006). Integrative exposure assessment through classification and regression trees on bioaccumulation of metals, related sampling site characteristics and ecoregions. Ecological Informatics 1, 55-65.

Postel, S.L. (2000). Entering an era of water scarcity: the challenges ahead. Ecological Applications 10, 941–948.

Prati, L., Pavanello, R., Pesarin, F. (1971). Assessment of Surface Water Quality by a Single Index of Pollution. Water Research 5, 741-751.

Quinlan, J.R. (1992). Learning with continuous Classes. In 5th Australian Joint Conference on Articial Intelligence, 343-348.

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

R Development Core Team (2009). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria.

Radif, A.A. (1999). Integrated water resources management (IWRM): an approach to face the challenges of the next century and to avert future crises. Desalination, 124, 145-153.

Ramírez, A., Viña, G. (1999). Colombian limnology. Contributions to its knowledge and statistical analysis. Univ. Jorge Tadeo Lozano - BP exploration, Bogotá.

Randin, C.F., Dirnbock, T., Dullinger, S., Zimmerman, N.E., Zappa, M., Guisan, A. (2006). Are niche-based species distribution models transferable in space? Journal of Biogeography 33, 1689–1703.

Rauch W., Henze M., Koncsos L., Reichert P., Shanahan P., Somlyódy L., Vanrolleghem. P. A. (1998). River Water Quality Modelling: I. State of the art. Water Science and Technology 38 (11), 237-244.

REBECCA (2004). Relationships between ecological and chemical status of surface waters- REBECCA project WP4 Rivers. Deliverable 6. Report on existing methods and relationships linking pressures, chemistry and biology in rivers. Chapter 1 Hydromorphological pressures and Chapter 3 Pressures from organic matter. Edited by Jens Møller Andersen, Michael Dunbar and Nikolai Friberg. Oslo, Norway

Refsgaard, J., van der Sluijs, J., Højberg, A., Vanrolleghem, P. (2007). Uncertainty in the environmental modelling process - A framework and guidance. Environmental Modelling & Software 22, 1543-1556.

Reichert, P. (2001). River Water Quality Model no. 1 (RWQM1): Case study II. Oxygen and nitrogen conversion processes in the River Glatt (Switzerland). Water Science and Technology 43, 51-60.

Reichert, P., Borchardt D., Henze M., Rauch W., Reichert P., Somlyódy L., Vanrolleghem P.A. (2001a). River Water Quality Model No.1: II. Biochemical process equations. Water Science and Technology 43 (5), 11-30.

Reichert, P., Borchardt, D., Henze, M., Rauch, W., Shanahan, P., Somlyódy, L., Vanrolleghem, P. A. (2001b). River Water Quality: Model No. 1 - IWA Scientific & Technical Reports No. 12. IWA Task Group on River Water Quality Modelling. IWA Publishing.

Rekolainen, S., Kämäri, J., Hiltunen, M. (2003). A conceptual framework for identifying the need and role of models in the implementation of the Water Framework Directive. Journal of River Basin Management 1 (4), 347–352.

Rushton, S.P., Ormerod, S.J., Kerby G. (2004). New paradigms for modelling species distributions? Journal of Applied Ecology 41, 193-200.

Schuwirth, N., Acuña, V., Reichert P. (2011). Development of a mechanistic model (ERIMO-I) for analyzing the temporal dynamics of the benthic community of an intermittent Mediterranean stream. Ecological Modelling 222, 91–104.

Segurado, P, Araujo, M.B. (2004). An evaluation of methods for modelling species distributions. Journal of Biogeography 31, 1555–1568.

Shanahan, P., Borchardt, D., Henze, M., Rauch, W., Reichert, P., Somlyody L., Vanrolleghem P.A. (2001). River water quality model no.1 (RWQM1): I. Modelling approach. Water Science and Technology 43 (5), 1-9.

Solomatine, D., See L.M., Abrahart, R.J. (2008). Data-Driven Modelling: Concepts, Approaches and Experiences. Chapter 2. In: Abrahart, R.J., See, L.M., Solomatine, D.P. (Eds.). Practical Hydroinformatics. Computational Intelligence and Technological Developments in Water Applications. Springer. p. 17-30.

Stalnaker B, Lamb BL, Henriksen J, Bovee K, Bartholow J. (1995). The Instream Flow Incremental Methodology: A Primer for IFIM. Biological Report 29. United States National Biological Service, Fort Collins, Colorado.

Statzner, B., Higler, B. (1986). Stream hydraulics as a major determinant of benthic invertebrate zonation patterns. Freshwater biology 16 (1): 127-139.

Statzner, B., Gore, J., Resh, V. (1988). Hydraulic stream ecology - Observed patterns and potential applications. Journal of the North American Benthological society, 7(4), 307-360.

Strauss, B., Biedermann, R. (2007). Evaluating temporal and spatial generality: how valid are species-habitat relationship models? Ecological Modelling 204, 104–114.

Stravs, L., Brilly, M., Sraj M. (2008). Precipitation interception modelling using machine learning methods – The Dragonja River Basin Case Study. In: Abrahart, R., See L.M., Solomatine D.P. (Eds), Practical Hydroinformatics, Springer Berlin 68, 347-358.

Symonds, M., Moussalli, A., 2011. Model selection, multimodel inference and model averaging using Akaike’s information criterion: an introduction for statistically terrified behavioural ecologists. Behavioral Ecology and Sociobiology 65, 13-21.

Timm, H., Kaeiro, K., Moels, T., Virro, T. (2011). An index to assess hydromorphological quality of Estonian surface waters based on macroinvertebrate taxonomic composition. Limnologica 41 (4): 398 - 410.

Tirelli, T., Pozzi, L., Pessani, D. (2009). Use of different approaches to model presence/absence of *Salmo marmoratus* in Piedmont (Northwestern Italy). Ecological Informatics 4, 234-242.

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

TULAS (2002). Legislation for environmental quality and effluent discharges: water resources. Unified text of the secondary environmental legislation in Ecuador, Book VI, Title IV, annex 1.

USEPA-United States Environmental Protection Agency (2011). A Primer on Using Biological Assessments to Support Water Quality Management. EPA 810-R-11-01. U.S. Environmental Protection Agency, Washington, DC.<

http://water.epa.gov/scitech/swguidance/standards/criteria/aqlife/biocriteria/upload/primer_update.pdf >. Accessed October 2011.

USGS-United States Geological Survey, Department of the Interior (2001). PHABSIM for Windows. User's Manual and Exercises Midcontinent Ecological Science Center. November 2001.Open File Report 01-340, 228 p.p.

Vandenberghe, V., van Griensven, A., Bauwens, W. (2005). Propagation of uncertainty in diffuse pollution into water quality predictions: application to the River Dender in Flanders. Water Science & Technology, 51 (3–4), 347–354.

van Griensven A, Breuer, L., Di Luzio, M., Vandenberghe, V., Goethals, P., Meixner, T., Arnold, J., Srinivasan R. (2006). Environmental and ecological hydroinformatics to support the implementation of the European Water Framework Directive for river basin management. Journal of Hydroinformatics 8 (4), 239–252.

van Griensven, A., Bauwens, W. (2001). Integral modelling of catchments. Water Science & Technology, 43(7), 321–328.

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

Vanrolleghem P.A., Fronteau C. and Bauwens W. (1996). Evaluation of design and operation of the sewage transport and treatment system by an EQO/EQS based analysis of the receiving water immission characteristics. In: Proceedings WEF Specialty Conference on Urban Wet Weather Pollution: Controlling sewer overflows and stormwater runoff, pp. 14.35.–14.46. Quebec City, Canada. June 16-19, 1996.

Vanrolleghem, P.A, Borchardt, D., Henze, M., Rauch, W., Reichert, P., Shanahan, P. and Somlyódy, L. (2001). River Water Quality Model no. 1 (RWQM1): III. Biochemical submodel selection.

Vanrolleghem, P.A. (2010a). Integrated assessment for Water Framework Directive implementation: data, economic and human dimension. Volume 2. In: Vanrolleghem P.A. (Ed.). IWA Publishing, London UK. p 305.

Vanrolleghem, P.A. (2010b). Modelling aspects of Water Framework Directive implementation. Volume 1. In: Vanrolleghem P.A. (Ed.). IWA Publishing, London UK. p 331.

Vannote R.L., Minshall G.W., Cummins K.W., Sedell J.R., Cushing C.E. (1980). The river continuum concept. Canadian Journal of Fisheries and Aquatic Sciences, 37, 130-137.

VARKOM, MOSTforWATER, Ghent University (2010). ECO-TEHPROJECT, Final Technical Report WATROPEC Project, Varaždin, Croatia.

Vaughan, I.P., Diamond, M., Gurnell, A.M., Hall, K.A., Jenkins, A., Milner, N.J., Naylor, L.A., Sear, D.A., Woodward, G., Ormerod, S.J. (2009). Integrating ecology with hydromorphology: a priority for river science and management. Aquatic Conservation: Marine and Freshwater Ecosystems 19, 113-125.

Vayssières, M., Plant, R., Allen-Diaz, B. (2000). Classification trees: An alternative non-parametric approach for predicting species distributions. Journal of Vegetation Science 11 (5), 679-694.

Verbyla, D.L., Litvaitis, J.A. (1989). Resampling methods for evaluating classification accuracy of wildlife habitat models. Environmental Management 13, 783–787.

Voinov, A., Bousquet, F. (2010). Modelling with stakeholders. Environmental Modelling & Software 25 (11), 1268-1281.

Voinov, A., Fitz, C., Boumans, R., Costanza, R. (2004). Modular ecosystem modelling. Environmental Modelling & Software 19, 285-304.

von der Ohe, P.C, De Deckere, E., PruB, A., Muñoz, I., Wolfram, G., Villagrasa, M., Ginebreda A, Hein M, Brack W. (2009). Toward an integrated assessment of the ecological and chemical status of european river basin. Integrated Environmental Assessment and Management 5 (1), 50-61.

Wagener, T., Lees, M.J., Wheater, H.S. (2004). A toolkit for the development and application of parsimonious hydrological models, in: Singh VP, Frevert D (Eds.) Mathematical models of large watershed hydrology. Water Resources Publications, pp 87–136.

Wang, Y., Witten, I.H. (1997). Induction of model trees for predicting continuous classes. In poster and papers of the 9th European Conference on Machine Learning. Springer.

Warmink, J.J., Janssen, J., Booij, M.J., Krol, M.S. (2010). Identification and classification of uncertainties in the application of environmental models. Environmental Modelling & Software 25, 1518-1527.

Whitehead, P.G., Young, P., Romberger, O. (1979). A systems model of stream flow and water quality in the Bedford-Ouse River I. Stream flow modelling. Water Research 13, 1155–1169.

Willems, W., Goethals, P.L.M., Eynde, Dries, V., Hoey, G.V., Lancker, V.V., Verfaillie, E., Vincx, M., Degraer, S. (2008). Where is the worm? Predictive modelling of the habitat preferences of the tube-building polychaete Lanice conchilega. Ecological Modelling 212, 74-79.

Willems P., Berlamont J. (2002). Probabilistic emission and immission modelling: case-study of the combined sewer-WWTP-receiving water system at Dessel (Belgium). Water Science and Technology 45 (3), 117–124.

Willmott, C.J. (1981). On the validation of models. Physical Geography 2, 184–194.

Witten, I.H., Frank, E., Hall, M.A., 2011. Data mining: practical machine learning tools and techniques, third edition, Morgan Kaufmann, San Francisco.

WWF-World Wildlife Fund (2003). Water and Wetland Index – Critical issues in water policy across Europe, Results overview for the Drava river basin (Croatia), Available at: http://www.wwf.org/.

Zúñiga, M.C., Cardona, W. (2009). Water quality and environmental flow bioindicators, (in Spanish). In: Cantera, J., Carvajal, Y., L., Castro, L. (Eds.) Environmental flow: Concepts, experiences and challenges, Ed. Del Valle University, Cali, Colombia, pp. 167-198.

Zuur, A.F., Ieno, E.N., Smith, G.M. (2007). Analysing Ecological Data. Springer, New York, U.S.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., Smith, G.M. (2009). Mixed Effects Models and Extensions in Ecology with R, first edition. Springer, New York, U.S.

Zuur, A.F., Ieno, E.N., Elphick C.S. (2010). A protocol for data exploration to avoid common statistical problems. Methods in Ecology and Evolution 1, 3–14.

Summary

Worldwide water managers invest large financial resources in infrastructures used for environmental protection, such as the collection and treatment of wastewater. However, quantifying a priori the effect of such investment programmes on the (ecological) river water quality is not straightforward. Modelling is an effective tool to investigate or to predict the ecological state of water resources and the response to natural driving variables or anthropogenic pressures. In developing countries and those which are in the process to join the European Union (EU), the impact of sanitation infrastructures (e.g. wastewater treatment plants (WWTP)) is typically assessed considering the achievement of legal physicochemical quality standards, but ignoring the ecological water quality of the receiving river. Natural systems are very complex with several processes occurring simultaneously and interacting. For instance, local conditions of current velocity, type of substratum and channel morphology influence the impact of physicochemical pollution on the river ecology. Thus, this traditional approach based only on physicochemical modelling for ecological protection or development of river restoration programs is not enough. Moreover, European (Water Framework Directive (WFD), 2000/60/CE) and American (Clean Water Act of 1972 and the Water Quality Act of 1987) legislation, changed the conventional practice by considering bioassessment and biocriteria in water resource assessment and management. Additionally, this legislation requests to use integrated approaches for decision making.

Current practice in model integration does not consider the simultaneous effect of hydromorphological disturbances and physicochemical pollution on the river ecology. Therefore, the aim of this study was to develop and to evaluate an integrated ecological modelling framework for decision support in river management. The proposed conceptual framework allows assessing ecological degradation in rivers and streams, helps to understand this problem and could provide crucial information for water managers in environmental decision making. This conceptual framework is called *Integrated Ecological Modelling Framework* (IEMF). This framework considers physicochemical pressures, such as the discharge of WWTP and hydromorphological pressures, such as changes in water course, current velocity, water depth, riverbed sediment composition and bank structure. Such comprehensive evaluation could not be achieved when looking at each individual component of the system separately (i.e. the impact of a WWTP effluent,

on the receiving river and a dam). The proposed IEMF was tested and validated on three case studies, in rivers with different geographical locations, altitude, size and pollutions problems: (1) a deep lowland river in a tropical region, the Cauca river in Colombia; (2) a shallow mountain river in a tropical region, the river Cuenca in the Andes of Ecuador; (3) a lowland river in a temperate zone, the Drava river in Croatia. The proposed research deals with the integration of river water quality and quantity models with two types of ecological models, river ecological assessment and species distribution models to predict the habitat suitability for selected species of macroinvertebrates. Moreover, a model assessing the WWTP processes was included in the IEMF considered for the Drava river.

The proposed model integration between WWTP, water quality, water quantity and river ecological assessment models is a feasible approach to evaluate the impact of sanitation infrastructure, such as WWTPs, on the ecological state of the receiving river. The IEMF can help to calculate the needed reductions in wastewater discharges of organic matter to meet biological water quality criteria. Potential investment scenarios of the wastewater treatment infrastructure (e.g. upgrading of the WWTP) in the three case studies were implemented and their impact on the ecological water quality of the receiving river were assessed. In general, it was found that the foreseen investments in sanitation infrastructure and current river restoration programs considered for the river basins in the three case studies are not enough to provide a good ecological water quality. Advanced investments, such as the collection and treatment of all domestic and industrial wastewater received by the rivers, the control and monitoring of the diffuse pollution sources, the treatment of the combined sewer overflows and the upgrading of the existing WWTP, with nitrogen and phosphorous removal are required.

The ecological models developed helped identifying that the impact of physicochemical pollution on the river ecology, generated by the discharge of wastewaters, is significantly influenced by local hydromorphological conditions. To this end, the IEMF considered for the hydromorphological assessment three elements: (1) average water depth; (2) average water velocity; (3) a variable called ‘Type’ that records information on the hydromorphological structure of the water body. Two categories or levels were defined for this Type variable: (1) hydromorphologically favourable: natural bank structure, mixed bottom substrate, thin sludge layer, meandering, heterogeneous bank and bottom structure;

and (2) hydromorphologically unfavourable: artificial bank structure, thick sludge layer, straight waterway, homogeneous bank and bottom structure.

It was found that species distribution models that predict the habitat suitability for selected species of macroinvertebrates, improved the understanding of the causal mechanisms and processes that affect the ecological water quality and shape macroinvertebrate communities in rivers. Simulations of pollution control scenarios implemented in the IEMF indicated an improvement in potential habitat availability for pollution sensitive taxa (e.g. Ephemeroptera and Trichoptera) and a decrease in potential habitat for pollution tolerant taxa (e.g. Haplotaxida and Physidae) as the pollution load from domestic and industrial wastewaters is reduced. The flexibility for updating or replacing the (ecological) models by better models when available, without having to change the IEMF, demonstrates the flexibility, applicability and transferability of this framework to other regions in the world. However, the main limitation of this approach is the availability of physicochemical, hydraulic and biological data that are collected simultaneously. Therefore, a change in the river monitoring strategy towards collection of data which include simultaneous measurements of variables is required to improve the ecological models.

Samenvatting

Wereldwijd investeren waterbeheerders grote sommen geld in de installatie van waterzuiveringsinfrastructuur. Op voorhand bepalen wat het effect is op de ecologische waterkwaliteit van zulke investeringen is echter niet evident. Het gebruik van modellen wordt aanzien als een efficiënte tool om de ecologische status van de waterkwaliteit te onderzoeken. In ontwikkelingslanden en landen die zich bij de EU aansluiten wordt het effect van waterzuiveringsinstallaties typisch onderzocht op basis van de kwaliteitsnormen die gehanteerd worden voor de fysicochemische parameters zonder hierbij rekening te houden met de ecologische waterkwaliteit. Deze traditionele aanpak op basis van fysicochemische modellen ter bescherming van de ecologie of het sturen van rivierherstelprojecten heeft verschillende tekortkomingen. Daarenboven hebben de Europese (Water Framework Directive (WFD), 2000/60/CE) en de Amerikaanse (Clean Water Act of 1972 and the Water Quality Act of 1987) wetgeving de algemene beoordelingscriteria veranderd en maken ze nu ook gebruik van biologische criteria bij het beoordelen van de waterkwaliteit en het beheer van water. Daarenboven stelt deze wetgeving dat er gebruik moet gemaakt worden van een geïntegreerde aanpak bij de besluitvorming.

Momenteel wordt er bij de integratie van modellen weinig of geen rekening gehouden met het gecombineerd effect van hydromorfologische en fysicochemische verstoringen op de ecologische waterkwaliteit. Het doel van deze studie is daarom om een geïntegreerd kader voor een ecologisch model te ontwikkelen en te evalueren ter ondersteuning van het rivierbeheer. Het voorgestelde conceptuele kader laat de beoordeling van de ecologische status van de rivier toe, voorziet oplossingen voor een slechte status en geeft cruciale informatie voor waterbeheerders. Dit conceptueel kader wordt het *Integrated Ecological Modelling Framework* (IEMF) genoemd. Dit kader neemt fysicochemische impacts zoals de lozing van het effluent van een waterzuiveringsinstallatie en hydromorfologische impacts zoals veranderingen in de stroomsnelheid gezamenlijk in beschouwing. Deze evaluatie kan nooit bereikt worden wanneer men de impact van elke component individueel gaat analyseren en beoordelen. Het voorgestelde IEMF werd getest en gevalideerd in drie verschillende case studies, in rivieren met een verschillende geografische ligging, hoogte grootte en pollutieproblemen: (1) een diepe laaglandrivier in een tropische regio, de Cauca rivier in Colombia; (2) een ondiepe bergrivier in een subtropische regio, de Cuenca rivier inde Andes in Ecuador; (3) een laaglandrivier in een mediterraan klimaat, de Drava rivier in Kroatië. Het voorgestelde onderzoek integreert waterkwaliteitsmodellen met waterkwantiteitsmodellen en twee types ecologische modellen: modellen om de waterkwaliteit te beoordelen en habitatgeschiktheidsmodellen. Daarenboven was een model om de processen die doorgaan in een waterzuiveringsinstallatie mee opgenomen in het DPPHER kader dat werd toegepast op de Drava rivier in Kroatië.

De modelintegratie van waterzuiveringsinstallatie, water kwaliteit, kwantiteit en ecologische beoordelingsmodellen vormt een goede aanpak om de impact van herstelmaatregelen te evalueren op de ecologische kwaliteit van de rivier. Het IEMF kan de nodige reductie in organisch materiaal en nutriënten bepalen dat nodig is om een goede waterkwaliteit te behalen. Potentiële investeringsscenario’s in waterzuiveringinstallaties (bv betere verwijdering van nutriënten in bestaande waterzuiveringsinstallaties) werden getest op drie verschillende case studies en de impact op de ecologische waterkwaliteit werd beoordeeld. In het algemeen werd er gevonden dat de huidige investeringen en rivierherstelprojecten onvoldoende zijn om een goede ecologische waterkwaliteit te behalen. Meer doorgedreven investeringen, zoals collecteren en behandelen van alle huishoudelijk en industrieel afvalwater, het beperken van diffuse puntlozingen, het vermijden van overstorten en het verhogen van de efficiëntie van bestaande waterzuiveringsinstallaties is noodzakelijk.

De ecologische modellen ontwikkeld in deze studie werden aangewend om aan te tonen dat de impact van fysicochemische polluenten afkomstig van de lozing van afvalwater op de ecologische rivierkwaliteit significant wordt beïnvloed door de heersende hydromorfologische condities. Het IEMF nam hierbij drie verschillende aspecten in beschouwing: (1) gemiddelde rivierdiepte; (2) gemiddelde stroomsnelheid en (3) de hydromorfologische structuur van de waterloop. Twee categorieën werden bepaald voor de hydromorfologische structuur: (1) hydromorfologisch gunstig (natuurlijke heterogene oeverstructuur, gemengd substraat, dunne sliblaag, meanderend) en (2) hydromorfologisch ongunstig (niet-natuurlijke oeverstructuur, dikke sliblaag, rechte waterloop, homogeen substraat).

Uit deze studie blijkt dat habitatgeschiktheidsmodellen voor een specifieke set van macroinvertebraten meer inzicht gaven in de onderliggende processen en verbanden die de ecologische waterkwaliteit bepalen en bepalend zijn voor het voorkomen van macroinvertebraten in rivieren. De habitatgeschiktheidsmodellen gaven een toename weer van het potentiële habitat voor pollutiegevoelige taxa (bv Ephemeroptera and Trichoptera) en een daling in het habitat van pollutietolerante taxa (bv Haplotaxida and Physidae) wanneer er een pollutieafname was als gevolg van een verbeterde waterzuivering.

De flexibiliteit met betrekking tot het updaten of vervangen van de bestaande modellen door andere modellen, zonder het IEMF te moeten veranderen geeft de flexibiliteit, toepasbaarheid en overdraagbaarheid van deze aanpak weer voornamelijk met betrekking tot het toepassen in andere werelddelen. De belangrijkste tekortkoming van het voorgesteld kader is de beschikbaarheid van hydraulische, fysicochemische en biotische data die gelijktijdig zijn verzameld. Daarom wordt het aanbevolen om de strategie aangaande de monitoring van rivieren aan te passen en er voor te zorgen dat er data wordt verzameld van alle verschillende variabelen op hetzelfde tijdstip om de ecologische modellen te kunnen optimaliseren.

CURRICULUM VITAE

IDENTIFICATION

Family names:	Holguin Gonzalez
First names:	Javier Ernesto
Birth date:	May 2, 1976
Place of birth	Cali (Colombia)
Gender:	Male
Civil State:	Single
Nationality:	Colombian
e-mail:	jaholgu@hotmail.com

UNIVERSITY STUDIES

2010-2013 Doctoral student in Applied Biological Science, Ghent University. Ghent, Belgium.
Thesis: Integrated ecological modelling for decision support in river management

2007-2009 Master of Science in Environmental Sanitation, Ghent University. Ghent, Belgium.
Thesis: Modelling the Ecological Impact of wastewaters on the Cauca river.

2001- 2003 Master of Science in Civil Engineering, with emphasis in Environmental Engineering. Los Andes University. Bogotá, Colombia.
Thesis: Determination of the re-aeration rates in a Colombian mountain river by using tracers.

1995-2001 Sanitary Engineering, Del Valle University. Cali, Colombia.
Thesis: Integrated solid waste management plan for the town La Victoria, Valle del Cauca.

PROFESSIONAL PROFILE

Specialist in the topic of Integrated Ecological Modelling for decision support in river management. Integration of water quality and hydraulic models with habitat suitability and ecological assessment models for decision support in river management. I have professional experience in activities related to water quality and hydraulic modelling and the design of engineering process for wastewater treatment plants. Additionally, I have worked in activities dealing with the promotion and execution of development projects, mainly related to water supply and sanitation systems.

PUBLICATIONS

A1-peer reviewed, Science Citation Index

Holguin-Gonzalez, J.E., Everaert, G., Boets, P., Galvis, A., Goethals, P.L.M. (2013). 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 and Software 48, 27–36.

Holguin-Gonzalez, J.E., Boets, P., Alvarado, A., Cisneros, F., Carrasco, M.C., Wyseure G., Nopens, I., Goethals, P.L.M (2013). 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., Boets, P., Everaert, G., Pauwels, I.S., Lock, K., Gobeyn, S., Benedetti, L., Amerlinck, Y., Nopens, I., Goethals, P.L.M. (submitted). Development and assessment of an integrated ecological modelling framework for decision support in water management. Water Resources Management.

Boets, P., Holguin-Gonzalez, J.E., Lock, K., Goethals, P.L.M. (2013). Data-driven habitat analysis of the Ponto-Caspian amphipod *Dikerogammarus villosus* in two invaded regions in Europe. Ecological Informatics 17, 36-45. doi:10.1016/j.ecoinf.2012.07.001

Papers of oral presentations at national and international conferences

Holguin J.E., Everaert G., Benedetti L., Amerlinck Y., Goethals P. (2012) Use of habitat suitability modeling in the integrated urban water system modeling of the Drava river (Varaždin, Croatia). In: H. Mader & J. Kraml (Eds.). Proceedings of the 9th International Symposium on Ecohydraulics-ISE 2012. Vienna, Austria, September 2012. ISBN: 978-3-200-02862-3. http://www.ise2012.boku.ac.at/papers/15613_2.pdf

Holguin J.E., Everaert G., Benedetti L., Amerlinck Y., Goethals P. (2012). Integrated ecological modelling for decision support in the integrated urban water system modelling of the Drava river (Varaždin, Croatia). In: R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.). International Environmental Modelling and Software Society (iEMSs), Sixth Biennial Meeting. Proceedings of the International Congress on Environmental Modelling and Software, "Managing Resources of a Limited Planet: Pathways and Visions under Uncertainty", Leipzig, Germany, July 2012. ISBN: 978-88-9035-742-8. http://www.iemss.org/sites/iemss2012/proceedings.html

Holguin J.E., Goethals P. L.M. (2010). Modelling the ecological impact of discharged urban waters upon receiving aquatic ecosystems. A tropical lowland river case study: city Cali and the Cauca river in Colombia. In: Swayne D, Yang W, Voinov A., Rizzoli A., Filatova T. (Eds.). Proceedings of the iEMSs Fifth Biennial Meeting: International Congress on Environmental Modelling and Software "Modelling for Environment's Sake". International Environmental Modelling and Software Society, Ottawa, Ontario, Canada, July 2010. http://www.iemss.org/iemss2010/Volume2.pdf, p. 1447-1455.

Holguin J.E., Goethals P.L.M., Galvis A. (2009). Modelling the ecological impact of wastewaters on the Cauca river (Colombia). Proceedings IWA Conference AGUA 2009. Integrated Water Resource Management and Climate Change. International Seminar: A New Paradigm on Integrated Water Management in Urban Areas. Universidad del Valle, Instituto CINARA. Cali, Colombia, 9 -13 November 2009.

Martínez A., Galvis A., Holguín J.E. (2007). Optimization of Cauca River Water Quality Modelling in the stretch La Balsa – Anacaro. Proceedings LATINOSAN 2007. The Latin American Conference on Sanitation. Thematic seminary of Prevention and Control of the Contamination of the Water Resource. 12 -16 of November of 2007. Cali, Colombia

Holguin J.E., Galvis A, Vélez C., Ramírez C. (2006). Modelling of environmental impact of discharged urban waters upon receiving aquatic ecosystems. Colombian case study: Cali city and the Cauca river. Proceedings 5th World Wide Workshop for Young Environmental Scientists WWW-YES 2006. Urban Waters: Resource or Risk. 9 – 12 de Mayo. Paris, France.

Holguin J.E., Vélez C., Galvis A., Ramírez C., Baena L., Duque A. (2005). Implementation of a dynamic model for the study of the water quality in the Cauca River. Proceedings IWA Conference Water 2005. International Seminary: Integrated management of services related to the water in nucleated establishments, 4 of November of 2005, Cali, Colombia.

Patiño P., Holguín J.E., Barba Ho L., Cruz C., Ramirez C., Duque A., Baena L. (2005). Methodology for the adaptation of a water quality index to the environmental conditions of the Cauca River in the stretch Salvajina-La Virginia. Proceedings IWA Conference Water 2005. International Seminary: Integral vision in the improvement of the Water Quality. 2-4 November of 2005, Cali, Colombia.

Cruz C., Barba Ho L., Holguín, J.E., Duque A., Patiño P. (2004). Methodological proposal for the identification of critical parameters of water quality in rivers, case of study Cauca River. Second International Environmental Congress of the Caribbean. CONCARIBE. Universidad Tecnologica de Bolivar. Cartagena de Indias. Colombia.

Holguín J.E., Camacho L.A. (2003) Determination of the re-aeration rate in a Colombian mountain river by means of the use of tracers. Proceedings IWA Conference Water 2003. International Seminary: The Hydroinformatic in the integrated water resource management. Cinara Institute, Del Valle University. Cartagena de Indias. Colombia. October 1 - 3 of 2003

Abstracts of oral presentations at national and international conferences

Holguin J.E., Everaert G., Goethals P. (2012). Use of multivariate statistics and machine learning techniques for integrated ecological modelling and decision support in river management. In De Baets B., Manderick B., Rademaker M., Waegeman W. (Eds.) Proceedings of the 21st Belgian-Dutch Conference on Machine Learning, BeneLearn 2012 and PMLS, Ghent, Belgium, 24-25 May 2012, p. 67.

Holguin J.E., Goethals P., Benedetti L., Amerlinck Y., Van Der Steede D. (2011). Use of multivariate statistics and machine learning techniques for ecological modeling, in the

integrated urban water system modeling of the Drava river (Varaždin, Croatia). In Jordán, F., Scotti, M., Lencioni V., (Eds.) Book of Abstracts of the 7th European Conference on Ecological Modelling (7th ECEM), Ecological hierarchy from the genes to the biosphere, Riva del Garda, Italy, 30 May–2 June 2011, p. 62.

Holguin J.E., Benedetti L., Amerlinck Y., Goethals P., Van der Steede D. (2010). Integrated urban water system modelling of the Drava river (Varaždin, Croatia) for cost-efficient wastewater treatment selection to meet the requirements of the European Water Framework Directive. In Goethals, P. (ed.) Proceedings of the 7th International Conference on Ecological Informatics (ISEI7), held in Ghent, Belgium, 13-16 December 2010. Ghent University Press, abstract for oral presentation, session-AS6, p. 133-134.

Holguin J.E., Boets P., Lock K., Goethals P. (2010). Habitat analysis of invasive crustaceans based on data-driven approaches applied on recently and long-term colonized habitats. In Goethals, P. (ed.) Proceedings of the 7th International Conference on Ecological Informatics (ISEI7), held in Ghent, Belgium, 13-16 December 2010. Ghent University Press, abstract for oral presentation, session-IS7, p. 198.

Holguin J.E., Alvarado A., Nopens I., Goethals P. (2010). Integrating hydrodynamic, physical-chemical and ecological models for decision support in water management of the Cuenca river in Ecuador. In Goethals, P. (ed.) Proceedings of the 7th International Conference on Ecological Informatics (ISEI7), held in Ghent, Belgium, 13-16 December 2010. Ghent University Press, abstract for oral presentation, session-DSS7, p. 176-177.

Paz Cortez Y., Holguin J.E., Galvis A., Goethals P. (2010). Integrated and model-based ecological assessment of the Cauca river (Colombia). In Goethals, P. (ed.) Proceedings of the 7th International Conference on Ecological Informatics (ISEI7), held in Ghent, Belgium, 13-16 December 2010. Ghent University Press, abstract for oral presentation, session-AS3, p. 130.

Abstracts of poster presentations at national and international conferences

Holguin J.E., Goethals P.L.M., Galvis A. (2010). Integrated ecological modelling for decision support in river management. A lowland river case study (Cauca river in Colombia). In: 16th PhD Symposium on Applied Biological Sciences. Faculty of Bioscience Engineering, Ghent University. Held in Ghent, Belgium, 20 December 2010. Ghent University and Katholieke Universiteit Leuven Press, abstract for poster presentation, session E2. P. 53

Chapters in books

Martinez A., Galvis A., Holguin J.E. (2008). Optimization of the Cauca River Water Quality Modelling in the stretch La Balsa – Anacaro. Section: La Balsa-Anacaro. In: Basic and environmental sanitation in Latin America (in Spanish). LITOCENCOA Ed. Colombia, ISBN: 978-958-44-3433-3, pp. 289 – 300.

Holguin J.E., Velez C., Galvis A., Ramírez C., Baena L., Duque A. (2007). Implementation of a dynamic model for the study of the water quality in the Cauca River. In: Advances in research and development in water and sanitation for meeting the Millennium Development Goals (in Spanish), Restrepo I, Sánchez L.D., Galvis A., Rojas

J., Sanabria I.J. (Eds.) Del Valle University Editorial Program. Colombia, ISBN: 95-86706-08-7, pp. 87-96.

Galvis A., Holguin J.E., Ramírez C., Velez C., Baena L., Duque A. (2007). Water quality of the Cauca river and its tributaries - Chapter 7. In: The Cauca river at its high Valley: a contribution to the knowledge of one of the most important Colombian rivers (in Spanish), Ramírez C., Sandoval M. C. (Eds.) Del Valle University Editorial Program and CVC. Colombia. ISBN: 978-958-8332-10-9, pp 207-266.

Galvis A., Holguin J.E., Ramírez C., Velez C., Baena L., Duque A. (2007). Mathematical Modelling of the Cauca river - Chapter 8 Item 8.6, Water quality Modelling. In: The Cauca river at its high Valley: a contribution to the knowledge of one of the most important Colombian rivers (in Spanish), Ramírez C., Sandoval M. C. (Eds.) Del Valle University Editorial Program and CVC. Colombia. ISBN: 978-958-8332-10-9, pp. 309-323.

Patiño P., Holguín J.E., Barba Ho L., Cruz C., Ramirez C., Duque A., Baena L. (2007). Methodology for the adaptation of a water quality index to the environmental conditions of the Cauca River in the stretch Salvajina-La Virginia. In: Advances in research and development in water and sanitation for meeting the Millennium Development Goals (in Spanish), Restrepo I, Sánchez L.D., Galvis A., Rojas J., Sanabria I.J. (Eds.) Del Valle University Editorial Program. Colombia, ISBN: 95-86706-08-7, pp. 391-399.

ORGANIZATION OF SCIENTIFIC MEETINGS OR CONFERENCES

Co-organiser of the 7th International Conference of the Ecological Informatics Society, Ghent, Belgium, 13 to 16 December 2010.

EDUCATIONAL ACTIVITIES

Courses at Universities (Autonoma de Occidente University, 2013):

- Environmental Modelling
- Adaptation and Mitigation to Climate Change

Practical exercises in courses at Universities (Ghent University, 2010-2012):

- Water quality management
- Biological monitoring of aquatic ecosystems
- Technology for integrated water management
- Environmental ecology
- Aquatic ecology
- Ecotecnology

Trainer in (integrated ecological) water quality modelling:

- Course in integrated ecological modelling for the workers of VARKOM Company at Varaždin, Croatia. December 2010.

- Course in Modelling to support the implementation of the European WFD. Habitat suitability modelling using Logistic regression to predict aquatic macroinvertebrates. In the International Congress of Ecological Informatics (ISEI7-2010). Ghent, Belgium.
- Trainer in river water quality modelling using the QUAL2K and QUAL2Kw models. CODECHOCO, Colombia. April 2010. Quibdo, Colombia.
- Trainer in river water quality modelling using the QUAL2K model. CRC, Colombia. June 2005 – September 2005. Popayan, Colombia.

Tutor of master theses

- Damanik A. Minar Naomi (2012 – 2013). Monitoring and modelling of the ecological water quality of the Cuenca river in Ecuador.
- Donoso Natalia (2011-2012). Integrated water system modelling to assess ecosystem services in the Cuenca river basin (Ecuador).
- Gobeyn Sacha (2011-2012). Integrated modelling of the multifunctional ecosystem of the Drava river (Croatia)
- Cisneros Salvatierra Janeth G. (2011-2012). Integrated ecological assessment of the Drava river (Croatia)
- Paz Cortes Yensy.I. (2010-2011). Integrated ecological modelling and assessment of the Cauca river (Colombia)

Scientific awards

- Doctoral Special Research Fund of Ghent University (BOF09/24J/092) in Belgium. (3 months in 2013)
- Doctoral Special Research Fund of Ghent University (BOF01/WI0/611) in Belgium. (2011 – 2012)
- Master fellowship from Ghent University (2008 – 2009)
- Master fellowship from the European Union Alfa Project (2007 – 2008)
- Master fellowship from the Los Andes University (2001 – 2003)
- Student paper commendation for an excellent student paper and presentation - one of the best 10 papers and presentations. iEMSs 2010 Conference Student Awards. International Environmental Modelling & Software Society (iEMSs- 2010). Ottawa, Canada. 7 of July 2010
- Huber Technology Prize 2010: Future Water. Certificate of appreciation one of the best 10 papers. Huber-Technology-Foundation. Berching, Germany. 16th September 2010.
- Graduation with distinction. Master of Science in Environmental Sanitation, Ghent University. Ghent, Belgium. September 2009

International research stays

Integrated ecological modelling and assessment of the Drava river (Croatia). Faculty of Bio-Science Engineering, Ghent University. Varaždin, Croatia, September 2011 - October 2011.

Project South-South mobility allowances. Flemish Interuniversity Council – University Development Cooperation VLIR-UOS and the Programs VLIR- IUC Cuenca and VLIR SRS. Cuenca, Ecuador, September 2009 - December 2009.