018530 - SWITCH

Sustainable Water Management in the City of the Future

Integrated Project
Global Change and Ecosystems

Modelling the ecological impact of wastewaters on the Cauca River (Colombia)

Due date of deliverable: August, 2009
Actual submission date: January, 2011

Start date of project: 1 February 2006 Duration: 60 months

Lead partner: UNIVALLE

Revision: final

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

UNIVERSITEIT GENT

FACULTY OF BIOSCIENCE ENGINEERING

Department of Applied Ecology and Environmental Biology
Laboratory of Environmental Toxicology and Aquatic Ecology
Prof. Dr. Eng. P. Goethals

Ghent, 28th January 2011

Dear Sir or Madam:
UNESCO – IHE
Institute for Water Education
SWITCH Project
Delft, The Netherlands

Reference: Support certificate to Master Thesis of the student Javier Holguín at Ghent University

By means of this letter we certify that the Master Thesis: "Modelling the ecological impact of wastewaters on the Cauca river (Colombia)" done by the student Javier Holguín for the degree of Master of Environmental Sanitation, was supervised by the Professor Prof. Dr. Eng. Peter Goethals of Ghent University and Co-supervised by the Eng. Alberto Galvis from Cinara Institute at Universidad del Valle.

This research was developed with the support of SWITCH Project, during the academic year 2008-2009 in the Faculty of Bioscience Engineering of Ghent University.

Yours Faithfully,

Prof. Dr. Eng. PETER GOETHALS

Eng. MSc JAVIER HOLGUÍN

Aquatic Ecology Research Unit
Department Applied Ecology and Environmental Biology
Ghent University
J. Plateaustraat 22, B-9000 Gent (Belgium)
Tel: +32 (0)9 264 37 68 Fax: +32 (0)9 264 41 99
E-mail: Peter.Goethals@UGent.be
Secretariat: Mrs. Sigrid Schuermans Tel: +32 (0)9 264 37 65

FACULTY OF BIOSCIENCE

ENGINEERING ACADEMIC YEAR 2008 –

2009

MODELLING THE ECOLOGICAL IMPACT OF WASTEWATERS ON THE CAUCA RIVER (COLOMBIA)

JAVIER ERNESTO HOLGUIN

GONZALEZ PROMOTERS:

Prof. Dr. ir. PETER
GOETHALS Prof. M.Sc.
ALBERTO GALVIS C

Master's dissertation submitted in partial fulfilment of the requirements for the degree of Master of Environmental Sanitation

CERTIFICATION AND DECLARATION

I, JAVIER ERNESTO HOLGUIN GONZALEZ, declare that this is the result of my own work and that no previous submission for a degree has been made here or elsewhere. Works by others, which served as sources of information, have been duly acknowledged by references to the authors.

The author and the promoters give the authorisation to consult and to copy parts of this work for personal use only. Any other use is under the limitation of copyrights laws; specifically it is obligatory to specify the source when using results from this thesis after having obtained the written permission.

This research was performed at:

Laboratory for Environmental Toxicology and Aquatic Ecology

Department Applied Ecology and Environmental Biology

Faculty of Bio-engineering Sciences, Ghent University

J. Plateaustraat 22, B-9000 Gent (Belgium)

Tel. 0032 (0)9 264 37 65. Fax. 0032 (0)9 264 41 99

……………………………..	……………………………..
16/08/2009	16/08/2009
Prof. Dr. ir. Peter Goethals	Javier Ernesto Holguin Gonzalez
(Promoter)	(Master thesis author) Email:
peter.goethals@UGent.be	Email: Jaholgu@hotmail.com

SUMMARY

The Cauca river is one of most severe cases of contamination for domestic and industrial wastewater discharges in Colombia. The rapid urbanization and major economic development in the Cauca river´s geographical valley has led to dramatic degradation of the environment and increased health risks due to inefficient processing of the increased pollutant load effluents and solid wastes. This river during the year 2005 received in average 195 tons of organic matter load per day in terms of BOD_5 in the study zone. The high discharge of organic matter causes a high oxygen demand in the degradation process. Thus, one of the most sensitive problems in the Cauca river is the decrease of dissolved oxygen (DO) with concentrations near to zero (0) mg/l in some monitoring stations especially during dry season (low flows conditions). Low DO levels affect the ecosystem equilibrium and the functioning and survival of biological communities. For this reason, the main objective of this research is to contribute to the integrated water quality management of the Cauca river in Colombia, developing a mathematical model to investigate the ecological quality of this river under actual conditions as well as after different restoration actions.

The approach followed was to build statistical models that allow predicting the occurrence (multiple logistic regression models - MLRMs) and the abundance of macroinvertebrates (quasi-Poisson regression models - QPRMs) in this river under different conditions. Afterwards, an integration of these ecological models with the hydrodynamic and physical-chemical water quality model MIKE11 was performed. Finally, applications of the integrated ecological modelling were made for predicting the ecological impact of the scenarios for pollution control in the Cauca river's basin.

The assessment of the MLRMs reliability showed that the models for Ephemeroptera (AUC=1), Trichoptera (AUC=1), and Haplotaxida (AUC=0.926) correctly discriminates between occupied (presence) and unoccupied (absence) sites in the dataset. Regarding the predictive validation procedure for QPRMs, it was found that in general the models reproduce with good precision the tendencies and the maximum and minimum values of abundance data for each macroinvertebrate (i.e. Ephemeroptera, Trichoptera and Haplotaxida) and the BMWP index, with high R^2 values ($0.866 < R^2 < 0.998$).

The application of the integrated ecological modelling of the Cauca river showed that the MLRMs and QPRMs predict well the ecological impact of the scenarios for pollution control in the Cauca river's basin. Thus, in the scenario with the highest pollution reduction an improvement of the water quality of the Cauca river is achieved, which is represented with the presence and/or an increase of the number of pollution sensitive benthos (i.e. Ephemeroptera and Trichoptera) and the absence and/or a decrease of the number of pollution tolerant benthos (i.e. Haplotaxida). On the other hand, if the worst pollution condition scenario is considered a deterioration of the water quality is obtained, which is represented with the absence and/or a decrease of the number of pollution sensitive benthos and the presence and/or an increase of the number of pollution tolerant benthos.

The integrated ecological model proposed in this research is a powerful operational tool, which allows to model and to assess the ecological impact of wastewater discharges into the Cauca river and can help to calculate the needed reductions in wastewater discharges of organic matter to meet biological quality criteria in this river.

ACKNOWLEDGEMENTS

First of all I would like to thank my promoters, Prof. Dr. ir. Peter Goethals at Ghent University in Belgium and Prof. M.Sc. Alberto Galvis at Universidad del Valle in Colombia, 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)" and the Program "Master Grants For Candidates From Developing Countries" at Ghent University, for their support of my studies in the Master program in Environmental Sanitation and my thesis research.

In Colombia, I would like to thanks to the Cinara Institute for constantly inspiring me to perform an excellent job. Also in Colombia, I would like to thanks to the environmental authority Corporación Autónoma Regional del Valle del Cauca (CVC), the Natural Resources and Environmental Engineering School (EIDENAR) of the Universidad del Valle, the Cauca River Modelling Project (PMC) and the Cinara Institute, 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 PMC Project where I learn to love models.

Finally, I would like to thank my wonderful parents and my girlfriend Clara for their love and their unconditional support during these two long years, who always believed in and supported my choices in life. I cannot forget to say thanks to my friends and colleagues in Ghent, specially my Latin friends with whom I share the huge pile of extra-ecological knowledge, in particular related to culture, gastronomy, religion, psychology and politics. Thank you, guys for your company, good ideas and for your enormous support.

Javier

Ghent, 20th August 2009.

LIST OF ABBREVIATIONS

AIC : Akaike's information criterion

ANN: Artificial neural networks

AUC: area under the curve

BBN: Bayesian Belief Networks

BMWP: Biological Monitoring Working Party

BOD_5: Biological oxygen demand

CCA: Canonical Correspondence Analysis

COD: Chemical oxygen demand

CVC : environmental authority Corporación Autónoma Regional del Valle del Cauca

DHI: Danish Hydraulic Institute

DO: Dissolved oxygen

EQI: Environmental Quality Index

EQR: Ecological Quality Ratio

GAMs: Generalized Additive Models

GLMs : Generalized Linear Models

MLR: Multiple Linear Regression

MLRG: multiple logistic regression model

MSM: Multivariate Statistical Methods

N : Total Nitrogen

P : Total Phosphorous

PCA : Principal Component Regression

PMC : Cauca River Modelling Project

QPRM : quasi-Poisson regression models

LIST OF ABBREVIATIONS (Cont.)

RD: Residual Deviance

RDA: Redundancy Analysis

ROC curve: Receiver Operating Characteristics

RR: Ridge regression

SLRM: simple logistic regression model

SM: Statistical models

TSS: Total suspended solids

Univalle: Universidad del Valle

WFD: European Water Framework Directive

WFD-Explorer: Water Framework Directive Explorer

WPIs: Water Pollution Indices

WQ: water quality

WQIs: Water Quality Indices

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Cont.)

LIST OF PHOTOS

1. INTRODUCTION

1.1 PROBLEM DEFINITION

One of the worldwide problems that affect the quality of water resources, has been their use such as receiving aquatic ecosystems of controlled or uncontrolled discharges of wastes from agricultural, urban or industrial activities. These discharges can potentially affect human health and aquatic life, limit water uses, affect riverine ecology and cause loss of amenity. Furthermore, scarcity and misuse of fresh water pose a serious and growing threat to sustainable development and protection of the environment. 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. The optimal balance between the different stakeholder activities needs a very deep insight in the integrated water management. In this context, models can show the limitations of the self-cleaning capacity of water resources.

Indeed, water quality modelling is an effective tool to investigate the ecological quality of surface water resources. Nevertheless, until now ecological models have rarely been used to support river management and water policy. According to Goethals (2005), models have several interesting applications in this context. Firstly, through these models a better interpretation of the river status can be possible, the causes of the status 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.

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. Among the biological communities, the macroinvertebrates are by far the most frequently used group for bioindication 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).

The Cauca river is one of most severe cases of contamination for domestic and industrial wastewater discharges in Colombia. The main urbanization center and source of pollution that affects this river is the city of Cali, with more than two million inhabitants and limitations of the treatment of its wastewaters. This city discharged during the year 2005 around 75 tons of BOD_5 per day, which is around 38% of the total of wastewater discharged load of BOD_5 per day in the Valle del Cauca department. The sewer system of Cali has limitations in the operation and only 40% of the total flow of the wastewaters generated by the city is treated by primary treatment (with BOD_5 removal of 30%). The rest of the wastewater generated by the city does not receive any type of treatment and it is discharged to the Cauca river. These discharges of wastewater are producing an increasing deterioration of the water quality of the Cauca river. This pollution problem is critical after the river crosses the city of Cali, especially during dry season (low flows), when pollution can reach values of 7.5 mg/l of BDO_5, concentrations of Dissolved Oxygen (DO) near to zero (0) mg/l, values of Faecal and Total Coliforms in the order of $2{,}4 \times 10^8$ NMP/100ml and critical values for some heavy metals (CVC and Univalle, 2004a and 2007a; EMCALI and Univalle, 2006; and EMCALI, 2007).

A part of the discharges of the sewer system of the city (south drainage system) is located upstream of the water intake for the main drinking water plant of the city, which supplies nearly 75% of the population of Cali. Apart from collecting and discharging wastewater generated from the city, the south drainage system transports also leachate generated in the solid waste disposal site of the city called Navarro through the south channel. According to EMCALI (2007) and INGESAM (2005) there is evidence that there are hazardous substances present in the wastewater coming from the south drainage systems such as: phenols, chromium, mercury and lead coming from the industrial discharges to the sewage system (industries that produce medical and dental equipment, rubber, plastic, press materials and metals) and from leachate. The consequences of this pollution of the main water source for the city of Cali are reflected in the increase of the risk of using this water and an increment in technical requirements and cost for the water treatment.

The main environmental authority in the Cauca river´s geographical valley is the Corporación Autónoma Regional del Valle del Cauca (CVC), which has been working in the water quality monitoring program of the Cauca river and its tributaries since the 1970s. The main physical-chemical and bacteriological parameters are monitored four (4) times

per year. These data, enriched along the years, have supported the first simulations of quality of the water. Since 1972 the mathematical modelling approach has been used to support the generation of policy, plans and projects for the water quality improvement of the Cauca river and the control of wastewater discharges. During this process, limitations in the knowledge of software and also in the information required in terms of water quantity and quality have been faced. During the last decade (1997-2007) in the framework of the Cauca River Modelling Project (PMC), the water quality model software MIKE11 was used to simulate the hydrodynamic and water quality of the river. This modelling approach allowed a very deep insight of the processes that occur in the river under dynamic conditions, such as temporary variations of flows and polluting loads.

However, this water quality modelling just considered physical-chemical parameters, therefore biological components of the aquatic ecosystem were not taken into account. For this reason, in order to have a robust, reliable and effective tool to support river management and water policy in the Cauca river, it is necessary to develop and to apply an ecological modelling approach that integrates hydrodynamic, physical-chemical, and biological components.

1.2 OBJECTIVES AND SCOPE

The general objective of this research is to contribute to the integrated water quality management of the Cauca river in Colombia.

Specific Objectives

To develop a mathematical model to investigate the ecological quality of the Cauca river under actual conditions as well as after different restoration actions.

To integrate the hydrodynamic and physical-chemical water quality model with an ecological predictive model developed for the Cauca river.

Scope

The scope of this research is to gain insight of the ecological impact of wastewaters on the Cauca river. The proposed methodology aims to develop an integrated water quality and ecosystem modelling, using the mathematical simulation model available in the study area for the Cauca river (model MIKE11) and an ecological predictive model (mainly focuses on macroinvertebrates) developed for this river in the framework of this research.

The chapters of the thesis are arranged as follows: Chapter 2 presents a review of the state of the art of integrated ecological modelling of rivers, Chapter 3 gives a description of the materials and methods used in this research, Chapter 4 contains the results of this investigation, Chapter 5 presents a general discussion of the results and further research, and finally Chapters 6 and 7 contain the conclusions and references.

2. STATE OF THE ART OF INTEGRATED ECOLOGICAL MODELLING OF RIVERS

2.1 RIVER QUALITY ASSESSMENT

2.1.1 Physical-chemical water quality indices

2.1.1.1 Introduction

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, different organizations of several nationalities involved in water resources control have used a regular form of physical and chemical indices for water quality assessment. This has been more evident in the last decade of the 20th Century. In that decade the application of water quality indices was given important acceptance, which is made evident at the present time by an appreciable number of formulated indices in different countries around the world, from general to specific purposes (Fernandez and Solano, 2008).

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, thereby leading to evaluating resource by means of a mathematical expression representing all evaluated parameters. WQIs and WPIs reduce a great amount of parameters 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 parameters, 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.

2.1.1.2 Some examples of WQIs and WPIs of worldwide importance

The water quality index developed in 1970 by the National Sanitation Foundation (WQI-NSF) of the United States, is one of the most worldwide WQI used. The WQI-NSF has

been adapted for specific regions or countries. For instance the Oregon Water Quality Index (OWQI) developed for rivers in Oregon state in the United Sates (ODEQ, 1980), the ICA-CETESB (CETESB, 2002) applied in rivers in Sao Pablo state in Brazil and the ICAUCA (CVC and Univalle, 2004b) developed for the Cauca river in Colombia. In general the final result of WQIs is a value between zero (0) and one hundred (100), which can be expressed as well as a verbal description (e.g. a value of one hundred (100) for very good quality and a value of zero (0) for very bad water quality), a symbol or a colour (e.g. blue for excellent quality and red for very bad water quality).

Regarding the Water Pollution Indices (WPIs), in the Netherlands in the framework of the AMOEBA project (A General Method of Ecological and Biological Assessment) (Brink *et al.*, 1991) the following indices were developed: 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). In Colombia, Ramirez and Viña (1999) in the framework of some river studies related with the oil industry developed six (6) WPIs that allow a desegregated evaluation of water quality. These WPIs are the Organic Matter WPI (ICOMO) which evaluates BOD_5, total coliforms and dissolved oxygen; the Mineralization WPI (ICOMI) which evaluates hardness, conductivity and alkalinity; the Suspended Solids WPI (ICOSUS); the Eutrophication WPI (ICOTRO) which evaluates total phosphorus; the temperature WPI (ICOTEMP) and the pH WPI (ICOpH). In general, the final result of WPIs is a value between zero (0) for very low pollution level and one (1) for very high pollution level. These WPIs allow the study of particular pollution problems and avoid the inconvenience that certain environmental pollution variables remain hidden by other variables, which is observed as a whole in general water quality evaluation strategies (WQIs) (Fernández and Solano, 2008).

2.1.2 Ecological indices

2.1.2.1 Importance of biological monitoring and assessment

Monitoring the quality of a freshwater ecosystem should not rely on physical-chemical analyses alone. Biological monitoring and biological criteria provide the most robust approach to track the status of waters, because waterways that cannot support healthy

biological communities are unlikely to support human society for long (Gabriels, 2007). Biological monitoring can provide more information on the state of an ecosystem than physical-chemical monitoring 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. In certain cases biological communities already respond before analytical detection allows for. For these reasons, physical-chemical and biological monitoring should be considered as complementary instruments for ecological monitoring.

2.1.2.2 Advantages and disadvantages of monitoring and assessment methods based on macroinvertebrates

Among the biological communities, the macroinvertebrates are by far the most frequently used group for assessment of freshwater quality and bioindication in standard water management. 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). Macroinvertebrates are ubiquitous and abundant throughout the whole river system and they play an essential role in the functioning of the river continuum food web (De Pauw *et al.*, 2006).

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), so they do not have to be sampled very frequently (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 physical-chemical 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 monitors of river (water) quality also has limitations. Quantitative sampling for example is difficult because of their non-random distribution in the river bed (Goethals, 2005). 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. An appreciation of this seasonality enables this to be taken into account in 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. Of these the related factors of current velocity and nature of the substratum are overriding ones determining the nature of the community, especially in relation to invertebrates. Since these factors 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 benthic conditions are selected or a typology is developed consisting of distinct river types with adapted sampling and assessment systems (Goethals, 2005).

A last limitation of macroinvertebrates is their restricted geographic distribution, the incidence and 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).

2.1.2.3 Different assessment approaches based on macroinvertebrates

Analysis of the macroinvertebrate communities in rivers can theoretically be structural, functional, taxonomical and non-taxonomical in approach (Goethals, 2005). Most of the actually used bioassessment systems are, however, structural and taxonomical, which means relying, for example, on the presence or absence of particular taxa, the sensitivity of particular taxa, the taxa richness, taxa abundance and taxa diversity. All of this information

can be converted into numerical values, including indices and scores. Most assessment methods are based on the analysis of species assemblages or populations of particular taxonomic groups of benthic macroinvertebrates (e.g. oligochaetes, chironomids) (De Pauw *et al.*, 2006). Assessment methods based on organism-level indicators (biochemical, physiological, morphological deformities, behavioural responses and life-history responses) are not considered here.

2.1.2.3.1 Saprobic approach

Historically, the saprobic approach was the first biological river assessment system ever developed (Kolkwitz and Marsson, 1902). This method makes a classification based on organic pollution resistance of present species. However, it is difficult to determine values for each species regarding 'pollution sensitivity' and it needs identification up to species level (of course the method can also be applied on a lower identification method). The method moreover needs abundance values for each species (Goethals, 2007).

2.1.2.3.2 Diversity approach

The diversity approach uses three components of community structure: richness (B), evenness (D) and abundance (C). The method is based on the fact that stress reduces the biodiversity. Typical examples that have frequently been applied using macroinvertebrate communities are the Shannon-Wiener index (Shannon and Weaver, 1949), the Simpson index (Simpson, 1949), Brillouin's diversity index (Brillouin, 1951), the Margalef index (Margalef, 1958) and the Evenness index (Hill, 1973) (Gabriels, 2007). The objective aims at evaluating the community structure with respect to occurrence of species. The diversity indices relate the number of observed species (richness) to the number of individuals (abundance). The advantages of diversity indices lie in the fact that they are easy to use and calculate, applicable to all kinds of watercourses, have no geographical limitations and are best used for comparative purposes (De Pauw *et al.*, 2006). Major disadvantages are related to the fact that it is difficult to define reference values (Goethals, 2007).

2.1.2.3.3 Biotic approach

The biotic approach incorporates desirable features of the saprobic and diversity approaches by combining a quantitative measure of species diversity with qualitative information on the ecological sensitivities of individual taxa into a single numerical expression (De Pauw *et al.,* 2006). Two different types can be distinguished within the biotic approach: the table-based biotic indices and the formula-based biotic indices, often referred to as biotic scores. The objective of biotic indices or scores is to assess the biological water quality of running waters, in most cases based on macroinvertebrates, and to measure various types of environmental stress, organic waters, acid waters, etc (Goethals, 2005). The principle is that macroinvertebrate groups disappear as pollution increases and that the number of taxonomic groups is reduced as pollution increases. The advantages are that only qualitative sampling is required and that identification is mostly at family or genus level and that there is no need to count abundances per taxon. The problems on the other hand are how to determine representative reference communities to which the investigated stations can be compared to (Goethals, 2005).

Biotic indices (table-based): In the biotic index approach the index is deduced from a table that takes into account the number of taxa and the sensitivity of the most sensitive taxon encountered. The first index of this type was the Trent Biotic Index (Woodiwiss, 1964), later extended to an Extended Biotic Index (EBI; Woodiwiss, 1978). An example of biotic indices is the Belgian Biotic Index (BBI; De Pauw and Vanhooren, 1983) (Gabriels, 2007).

Biotic indices (formula-based): In the biotic score system a predefined score is allocated to each taxon. These individual taxon-scores depend on their sensitivity to pollution. For calculating the score of a site, all individual taxon-scores of the encountered taxa are summed. The biotic score may also include a measure of abundance of the organisms (Goethals, 2005). The best-known example of a biotic score is the BMWP (Biological Monitoring Working Party) score (Chesters, 1980) and its revised version (National Water Council, 1981) (Gabriels, 2007). Adaptations of the BMWP were formulated for the Iberian Peninsula (IBMWP; Alba-Tercedor *et al*., 2002), Hungary (MMCP; Csányi, 1998), Poland (BMWP-PL; Kownacki *et al*., 2004) (Gabriels, 2007).

Some attempts to assess the performance of biological assessment methods applied on developing countries have been done in Latin America based on macroinvertebrates, such as Nicaragua (Fenoglio *et al.*, 2002), Chile (Figueroa *et al.*, 2003) and Colombia (Gutierrez *et al.*, 2004a, b; Riss *et al.*, 2002; Roldán, 2003) (cited by Domínguez *et al.*, 2005) and Zúñiga, (2009). It is worth mentioning the adaptation of the BMWP for rivers located at the Colombian southwest region developed by Zúñiga (2009).

2.1.2.3.4 Multimetric approach

In multimetric systems, several metrics representing different characteristics of the community are combined into one index value or score which is an expression of the overall quality (Gabriels, 2007). It is assumed that incorporating more descriptors will result in an index being more diagnostic of ecosystem health. Multimetric systems may include structure metrics, community balance metrics, tolerance metrics, feeding group metrics and others (e.g. USEPA, 1996) (Goethals, 2005). An important advantage of multimetric indices is their flexibility. They can easily be adapted to a regional situation, by taking into account the most appropriate metrics and by evaluating each metric to an appropriate target. The flexibility of this type of index is probably the reason why the majority of the indices that were developed in recent years were multimetric indices (Gabriels, 2007).

2.1.2.3.5 Ecological Quality Ratio approach

In this approach the assessment is based on the degree of similarity of the community composition to a pre-defined target community. This target, usually called the reference, can be based on actual samplings, expert knowledge, historical data or predictive models, or a combination of these (Gabriels, 2007). The most straightforward method of comparing an observed to a reference community is to calculate the proportion of the assessment value obtained with any index system between both communities. This is called the Ecological Quality Ratio (EQR) according to the Water Framework Directive (EU, 2000). This ratio is expressed as a numerical value between zero and one, where zero represents a very bad ecological status and one a very good ecological status (Wallin *et al*., 2003; cited by Gabriels, 2007).

An example of an EQR is the Environmental Quality Index (EQI) based on the “River Invertebrate Prediction and Classification System” (RIVPACS) developed in the UK (Wright *et al*., 1993; Wright, 2000; cited by Goethals, 2005). This system produces a site-specific prediction of the macroinvertebrate taxa that should be present under undisturbed conditions based on a number of physical-chemical features of the examined site (Gabriels, 2007). These predicted reference conditions can then be compared with the observed macroinvertebrate communities by calculating the EQI, which is in accordance with the WFD. The RIVPACS EQI can be calculated with different metrics or indices, for example the BMWP, the ASPT or the number of taxa (Wright, 2000; cited by Gabriels, 2007). Based on RIVPACS, other similar models have been developed in Australia (AUSRIVAS: ‘Australian River Assessment Scheme’, e.g. Davies, 2000; Smith *et al*., 1999; cited by Goethals, 2005) and Canada (BEAST: ‘Benthic Assessment of Sediment’, Reynoldson *et al*., 2000; cited by Goethals, 2005).

2.1.2.3.6 Multivariate approach

This method is based on multivariate statistics – data analysis. The basis for the multivariate approach is the similarity index (Sandin *et al*., 2000). The most commonly used similarity index is the Jaccard’s index (Jaccard, 1908 and Washington, 1984; cited by De Pauw *et al.*, 2006). This index expresses the percentage of species shared between two sites. Other examples are the percentage similarity index (Whittaker, 1952), Bray-Curtis dissimilarity index (Bray and Curtis, 1957), Sorensen index (Sorensen, 1948) and Euclidean of ecological distance (Williams, 1971; cited by De Pauw *et al.*, 2006). All these indices give an indication how much a biological community at each sampled site is similar to the median of all reference communities and are not resulting in an assessment class as such (Goethals, 2005).

Multivariate techniques are, since the 1990s, also commonly applied for the development of multimetric systems. The selection of the metrics is based on how complementary or explanatory these are. The complementary of score systems is necessary to guarantee that correlated metrics do not dominate the overall assessment, while the explanatory aspects are interesting for obtaining insight in the causes of deterioration. Since the new millennium, also a shift in use from multivariate statistical (classification, ordination, regression, clustering, etc., based on data distribution functions) to soft computing (based

on heuristic search methods, e.g., artificial neural networks, inductive logic programming, etc.) techniques has started. Major examples of assessment systems using multivariate approaches are RIVPACS and AusRivAS (Davies, 2000; cited by De Pauw *et al.*, 2006).

Presently, the most commonly applied indices in Europe are based on the saprobic and biotic approach (De Pauw *et al.*, 2006). Recently, however, as an incentive of the European Water Frame Directive, also (stressor-specific) multimetric systems, originating from the USA, are now being developed and introduced (De Pauw *et al.*, 2006). In contrast with the saprobic and biotic indices, which are solely based on a community structure analysis, the multimetric assessment systems may also include functional and non-taxonomic characteristics. For the assessments based on macroinvertebrates, also more and more use is being made of multivariate analysis which has the advantage to clearly link the biological communities to the river typology (Goethals, 2005). Other characteristics which have received attention during the last decade in assessments are the macroinvertebrate community structure related to the feeding strategy, migration or habitat use (e.g. Index of Trophic Completeness (ITC), Pavluk *et al*., 2000) and the use of key or target species (in how far does the species or taxa composition correspond with the expected composition of a particular type of surface water) (e.g. Lorenz *et al*., 2004; cited by Goethals, 2005).

Some examples of multivariate statistical approach and soft computing techniques applied on developing countries have been done in Latin America, such as Ecuador (Domínguez, 2007) and Colombia (Ramirez and Viña, 1999). It is worth mentioning the Biotic Index (ICOBIO) developed by Ramirez and Viña (1999) which evaluates the diversity or biological structure of communities by mean of Jaccard's similarity index (Jaccard, 1908) and Bray-Curtis dissimilarity index (Bray and Curtis, 1957).

2.2 MAJOR TYPES OF RIVER SYSTEM MODELS

Basically, there are two types of river system models called stochastic and deterministic. If the model contains elements of randomness it is called stochastic. 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). Thus, in a stochastic model the parameters are estimated in terms of statistical distributions and the model will generate a range of values (rather than a single one) as model output in the form

of a frequency distribution of e.g. pollutant concentration. If the model contains no elements of randomness or does not comprise uncertainty, the model output is a single value. This type of model called deterministic, assumes that the future response of the system is completely determined by knowledge of the present state and the future measured inputs. A deterministic model can be further described as mechanistic (white-box), grey-box and black-box model. Mechanistic (white-box) models are based on physical, biological and chemical laws, such as conservation of mass, momentum and energy, whereas the black-box (data driven) models are those models that are not based on any physical or biological laws; instead they are based on a data driven transfer function (e.g. Artificial Neural Networks-ANNs). If a model contains elements of both the white-box and the black-box model, the model is called grey-box (Deksissa, 2004). With regards 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, inputs variables and state variables may vary with time, and thus result in a time variable output.

2.2.1 Hydrodynamic and hydraulic modelling

Flow of water in a river is described by the continuity and momentum equations. The form of a hydrodynamic model depends on assumptions made on characterizing turbulence. For water quality modelling of rivers mostly the well-known, cross-sectionally integrated (1D) Saint Venant equations or approximations to these equations are used. 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):

$$\frac{\partial Q}{\partial x} + \frac{\partial A_{cross}}{\partial t} = q \qquad (2.1)$$

Momentum equation (momentum balance):

$$\frac{1}{A_{cross}}\frac{\partial Q}{\partial t} + \frac{1}{A_{cross}}\frac{\partial}{\partial x}\left(\frac{Q^2}{A_{cross}}\right) + g\frac{\partial h}{\partial x} - g\left(S_o - S_f\right) = 0 \qquad (2.2)$$

local acceleration — convective acceleration — pressure force — gravity force — friction force

Kinematic wave (gravity force, friction force)

Diffusive wave (pressure force, gravity force, friction force)

Dynamic wave

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

Depending upon whether the flow is steady or unsteady (dynamic water movement) and which simplifications are made, many different forms and approximations to the Saint Venant equations are known. Thus, for water quality studies often the equation of steady, gradually variable flow is employed (which may be further simplified to the so-called Manning equation as done in QUAL2E or QUAL2K models). Unsteady models include the kinematic, diffusive, and dynamic wave approaches, all based on the continuity and momentum equations, which can be solved for instance by software such as MIKE11 (DHI, 1999). The difference stems from simplifications of the latter: dynamic wave models solve the full equation, diffusive ones exclude the acceleration terms, while kinematic ones disregard also the pressure gradient term that is essential for the description of backwater effects (Rauch *et al*., 1998).

The hydrodynamic equations are generally solved by efficient finite difference methods. For water quality issues the acceleration terms in the momentum equation rarely play a significant role and the typical time scales are amplified by conversion processes. For these reasons, the diffusive wave approach is often a satisfactory approximation (Rauch *et al*., 1998).

2.2.2 Physical-chemical water quality modelling

The significance of different water quality processes varies depending on the case study considered. In general physical-chemical water quality modelling includes model kinetics and mass transfer processes. Kinetic processes such as: dissolution, hydrolysis, oxidation, nitrification, denitrification, photosynthesis, respiration, excretion and death. Mass transfer processes such as: reaeration, settling, sediment oxygen demand, sediment exchange, and sediment inorganic carbon flux (Chapra *et al.,* 2008).

2.2.2.1 MIKE11 model

MIKE11 is a unidimensional model developed by the Danish Hydraulic Institute (DHI). This model has the capacity to simulate flow, sediment transport, morphological processes, and water quality, under non-permanent flow conditions. It has a series of integrated modules: Hydrodynamic, Advection – Dispersion, Water Quality and Sediment Transport. The Hydrodynamic module is the core of the system. The hydrodynamic module resolves integrated mass conservation and movement quantity equations, named Saint Venant equations which describe the non-permanent flow in open channels (DHI, 1999).

The water quality (WQ) module of MIKE11 deals with the basic aspects of river water quality in areas influenced by human activities such as oxygen depletion and ammonia levels as a result of organic matter loading. The WQ module is coupled to the advection–dispersion (AD) module, which means that the WQ module deals with transforming processes of compounds in the river and the AD module is used to simulate the simultaneous transport process. The WQ module solves the system-coupled differential equations describing the physical, chemical and biological interactions in the river. The river water quality can be dealt with six different levels of detail: 1st BOD-DO relationships, 2nd BOD-DO relationships including exchange with organic matter from the riverbed; 3rd BOD-DO relationships including nitrification; 4th BOD-DO relationships including exchange with the riverbed and nitrification and denitrification; 5th BOD-DO relationships including immediate and delayed oxygen demand and exchange with the river bed, and 6th BOD-DO relationships including all the above mentioned processes.

2.2.3 Ecological modelling for predicting macroinvertebrates in rivers

The ecologic status of surface water mainly depends on the actual and historical immission characteristics (physical-chemical characteristics), hydrologic/hydraulic regime and morphologic characteristics (Figure 2.1). 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 physical-chemical and biological processes that occur in the surface water (Bauwens, 2009).

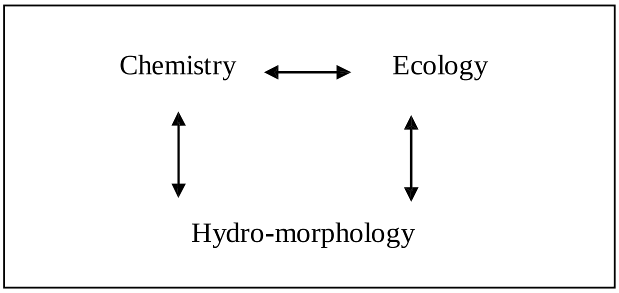


Figure 2.1 Driving forces affecting the ecologic status of the surface water

The use of appropriate mathematical models for surface water can help describe or predict ecological processes and response (at individual, population or community level) to natural driving variables or anthropogenic pressures (Figure 2.2). 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. Aquatic habitat models serve three main purposes: first, to predict the (probability of) occurrence, abundance or distribution of species based on relevant abiotic and biotic variables at new sites and/or future times, second to improve the understanding of species-habitat relationships and third, to quantify habitat requirements.

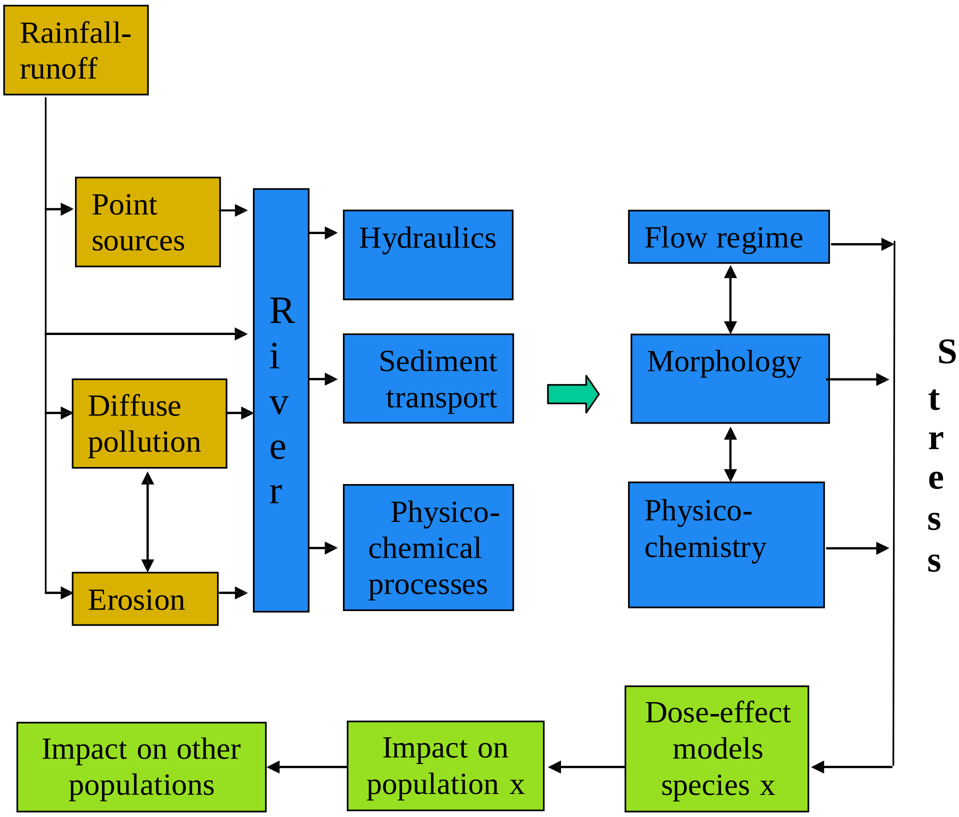


Figure 2.2 Ecological Modelling Flow Chart in a river. Source: Bauwens (2009)

In general, two approaches can be followed when performing (ecological) modelling: mechanistic (i.e. food-webs) (De Laender *et al.*, 2005; Jorgensen S.E., 2008; Jorgensen and Bendoricchio, 2001) and data driven modelling (e.g. statistical models (SD) or ANNs for habitat suitability modelling) (Ahmadi-Nedushan, 2006; Goethals, 2005; NIVA, 2007).). The first approach mathematically synthesizes available knowledge into a predictive framework, while the latter consists more of a data driven process. Which approach will be preferable in a specific case depends on the aim of the research, the knowledge of the ecological processes and state variables in the system, the required properties of the model and the data set available.

Examples of mechanistic models (i.e. food-webs) are bio-geo-chemical and bio-energetic dynamic/static models. These models are based on mass and energy conservation principles and generally apply differential equations to express the dynamics of the system, quantifying the growth of many different populations while accounting for ecological interactions between these populations. Change in state variables are expressed as the results of the ingoing minus the outgoing processes. The process equations are based usually on causality; but in principle can also be a result of a statistic analysis of data. Mechanistic models may vary in their level of representation, model components, and temporal representation. On the basis of the level of representation, the mechanistic models vary from very simple (lumped models) to very complex (distributed models). In the lumped models, several processes may be combined and expressed as one, whereas in the distributed model, the model attempts to represent every significant process.

The application of mechanistic models has been mainly focused on lentic ecosystems (lakes, ponds, reservoirs and wetlands) and the prediction of phytoplankton (Saito, 1999), zooplankton (Rodrigues *et al.*, 2007), macrophytes (Velez, 2006) and fish communities (Nestler *et al.*, 2005). De Laender (2007) presented a dynamic ecosystem model for assessing toxic effects of chemicals on lentic ecosystems considering interactions between species (phytoplankton, zooplankton, macrophytes, planktivorous and piscivorous fish). Examples of the application of mechanistic models for predicting macroinvertebrates in lotic ecosystems (rivers and streams) are rather limited and hardly described in literature.

Examples of data driven models used for predicting aquatic communities (habitat suitability models) in rivers on the basis of abiotic stream characteristics include Statistical Models (SM), classification trees, ANNs. Other methods such as fuzzy logic based models and Bayesian Belief Networks (BBNs) are called expert knowledge models, because they introduce the advantage of linguistic properties (fuzzy logic) and they have so called learning capability if large data sets are available. Data driven techniques are interesting when a lot of data of good quality are available.

The use of statistical models to predict (the likelihood of) occurrence, abundance or distribution of species based on relevant variables is becoming an increasingly important tool in conservation planning and wildlife management (Ahmadi-Nedushan, 2006). The purpose of the statistical model is to provide a mathematical basis for interpretation, examining such parameters as 'fit' (Do the measured predictors adequately explain the response?), 'strength' of association (Is the relationship between the response and the predictors significant?), and to ascertain the contributions and roles of the different variables (Guisan *et al.*, 2002). Classification trees extract simple rules from large quantities of data, while ANNs are able to establish patterns and characteristics in situations where rules are not known. Fuzzy logic and BBNs on the other hand allows to process unreliability and inaccuracy of data and to incorporate external expert knowledge, what is in particular useful to predict rare species, because the development of data driven models is not possible under data-poor conditions. However, also for this type of models the availability of proper and reliable expert knowledge is of crucial importance, as well as at least a good validation set of data (Goethals, 2005*).*

Applications of data driven methods to predict biological communities present in rivers, focused on macroinvertebrates, are described by Ahmadi-Nedushan *et al.* (2006), Goethals (2005) and Adriaenssens (2004). The application of these mathematical modelling techniques in rivers in tropical region countries in South America, is rather limited and hardly described in literature. An application of ANN and fuzzy logic networks to predict the biologic water quality in terms of the aquatic macroinvertebrates of rivers at the sabana of Bogota-Colombia is presented by Gutiérrez *et al.*, (2004a and 2004b).

2.2.3.1. Statistical methods

Statistical models in ecology can be used as explanatory and predictive models. In general, explanatory models seek to provide insights into the ecological processes that produce patterns. Often, these relationships are determined from statistical models that ascertain the strength of the statistical relationship between a response (e.g. macroinvertebrates species presence) and a suite of one or more explanatory variables (e.g. water velocity, substrate, dissolved oxygen concentration). In contrast, predictive models typically seek to provide the user with a statistical relationship between the response and a series of predictor variables (hereafter simply called the predictors) for use in predicting the probability of species occurrence or estimating numbers of organisms (abundance) at new sites and/or future times. These models often use variable reduction techniques in the analytical phase and have as their goal a model that predicts the ecological attribute(s) of interest from a restricted number of predictors. The concept of parsimony, that the simplest explanation is best, is inherent in such modelling efforts. The reduced model typically has lower variance, which will trade off with bias in optimizing prediction error (Guisan *et al.*, 2002).

A wide array of habitat statistical models has been developed to analyse habitat-species relationship. Ahmadi-Nedushan *et al.* (2006) presented a review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment, which include: Multivariate Statistical Methods (MSM), classical regression techniques (simple and multiple linear regressions); modern regression techniques such as Generalized Linear Models (GLMs) and Generalized Additive Models (GAMs); Ridge regression and Principal component regression among others.

Multivariate Statistical Methods (MSM): Generally, physical habitat is dependent on more than one variable (e.g. depth, velocity, substrate, cover) and several suitability indices must be combined to define a composite index. Multivariate approaches are more appropriate for the analysis of aquatic habitat as they inherently consider the interrelation and correlation structure of the environmental variables (Ahmadi-Nedushan, 2006).

Classical regression techniques: Linear regression is one of the oldest statistical techniques, and has long been used in biological research. Multiple linear regression (MLR) is one of the most commonly used methods to describe relationships between

dependent variables (e.g. species abundance) and independent variables (e.g. abiotic predictors). Regression analysis uses associations between independent and dependent variables to relate a response variable to a single (simple regression) or a combination (multiple regression) of environmental variables. The relationship between species responses and environmental variables can be described as:

$$Y = \beta_0 + X^T \beta + \varepsilon = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m + \varepsilon \qquad (2.3)$$

where, Y denotes the response variable (e.g. abundance), β_0 is a constant called the intercept, $X = (X_1, \ldots, X_m)$ is a vector of m predictor variables, $\beta = \beta_1, \ldots \beta_m$ is the vector of m regression coefficients (one for each predictor), and ε is the error. The error represents measurement error, as well as any variation unexplained by the linear model. Regression coefficients are generally estimated using Ordinary Least Squares (OLS) algorithm to minimize the differences between predicted and observed response. If the response variable does not have a linear relationship with a predictor, a transformed term of the environmental variable can be included in the model.

Although a powerful approach in situations when appropriately applied, linear regression is limited by three main assumptions: 1) the errors ε_i are assumed to be identically and independently distributed; this includes the assumption that the variance of response variable (Y) is constant across observations; 2) model errors ε_i are assumed to follow a normal (Gaussian) distribution; and 3) the regression function is linear in the predictors. Violation of assumption 1 constitutes a limitation to the application of most parametric statistical models, and is directly related to data sampling. Typically, many data in ecology are not Gaussian and do not have a constant variance. As an example, count data (e.g. number of individuals or species) follow a Poisson distribution (Jones *et al.*, 2002; Ver Hoef, and Boveng, 2007), and their variance is proportional to their mean (Guisan *et al.*, 2002; Ahmadi-Nedushan, 2006).

A common way of dealing with departures from assumptions 1 and 2 is to transform the response variable (Y) so that it meets the criteria of normality and constant variance (e.g. Z = Log Y; Z = Ln Y; Z = 1/Y). Violations of assumption 3 have traditionally been dealt with by augmenting the predictors with polynomial terms, interactions and other non-

linear transformations of the original predictors, leading to a model non-linear in the X_j but linear in the parameters (Guisan *et al.*, 2002). However, in some cases, it may be impossible for the same transformation to create normally distributed errors, to stabilize the variance, and to lead to a linear model. Modern regression methods like Generalized Linear Models (GLMs) or Generalized Additive Models (GAMs) are better suited to analyse these types of data (Ahmadi-Nedushan, 2006; Lindsey, 1997).

Modern regression techniques: Regression analyses have been broadly applied in ecology. An important statistical development of the last 30 years has been the advance in regression analysis provided by GLMs modelling (McCullogh and Nelder, 1989) and GAMs (Hastie and Tibshirani, 1990). Since their development, both approaches have been extensively applied in ecological research, as evidenced by the growing number of published papers incorporating these modern regression tools (Guisan *et al.*, 2002; Ahmadi-Nedushan, 2006). In GLMs, data may be assumed to be from several families of probability distributions, many of which better fit the non-normal error structures found in most ecological data, therefore GLMs present the advantage of dealing with non-normal environmental variables. Fields where the use of GLMs and GAMs has proven particularly useful are the modelling of the spatial distribution of species and communities (Guisan *et al.*, 2002) and the prediction of occurrence and abundance of species (Jones *et al.*, 2002; Ver Hoef, and Boveng, 2007). Manel *et al*. (2000) presented an example of the application of GLMs (Multiple Logistic Regression) for the prediction of presence and absence of macroinvertebrates in Himalayan rivers.

Generalized Linear Models (GLMs): GLMs are mathematical extensions of traditional linear models (generalization of Multiple Linear Regression) that allow the mean of a population to depend on a linear predictor through a nonlinear link function and allow the response probability distribution to be any member of an exponential family of distributions. These models do not force data into unnatural scales, and thereby allow for nonlinearity and non-constant variance structures in the data. GLMs are a more flexible family of regression models, which allow other distributions for the response variable. Many widely used statistical models are GLMs. These include classical linear models with normal errors, logistic and probit models for binary data, and log-linear models for multinomial data. Statistical models can be formulated as GLMs by the selection of an

appropriate link function and response probability distribution. Data may be assumed to be from several families of probability distributions (exponential family of distributions), including the normal, binomial, Poisson, negative binomial, or gamma distribution, many of which better fit the non-normal error structures of most ecological data (Guisan *et al.*, 2002). Thus, GLMs are more flexible and better suited for analyzing ecological relationships, which can be poorly represented by classical Gaussian distributions.

Logistic regression (Manel *et al.*, 2000; Pearce and Ferrier, 2000; Ahmadi-Nedushan, 2006; Rushton *et al*., 2004; Reineking & Schroderc, 2006), to predict the (probability of) occurrence, and Poisson regression, quasi-Poisson regression and negative binomial regression (Jones *et al.*, 2002; Ver Hoef, and Boveng, 2007; Guisan *et al.*, 2002; White and Bennetts, 1996; Dolan *et al.*, 2000; Sudhir and Krishna, 2007), for modelling abundance data, are the most popular GLMs for modelling species and their relationships with the environment. All GLMs are comprised of three components; a response variable distribution which identifies the response *Y*, a linear predictor which specifies the environment variables used as predictors in the model, and the link function *g*, which describes the functional relationship between the linear predictors and the expected value $E(Y)$ of the response variable. The GLM relates a function of the mean to environmental variables through a prediction equation of a linear form (Ahmadi-Nedushan, 2006). Using the same notations used for multiple linear regression (i.e. equation 2.3.) the following equation represents a GLM:

$$g(\;_{(X)}) \quad _0 \quad X^T \quad _0 \quad _1X_1 \quad _2X_2 \quad ... \quad _mX_m \qquad (2.4)$$

Generalized Additive Models (GAMs)**:** GAMs are non-parametric extensions of GLMs; the only underlying assumption made is that the functions are additive and that the components are smooth. While GLMs extend the application of the classical regression into other statistical distributions (binomial, Poisson, gamma, negative binomial), GAMs estimate response curves with non-parametric smoothing functions instead of parametric terms (e.g. $ax + bx^2$) (Ahmadi-Nedushan, 2006; Guisan *et al.*, 2002). A GAM, like a GLM, uses a link function to establish a relationship between the mean of the response variable and a 'smoothed' function of the predictor variables. The GAMs model approach provides more flexibility in the model form. It is actually possible to combine a standard GLM in some of the predictors with nonparametric regressions in the other (Ahmadi-Nedushan,

2006). The strength of GAMs is their ability to deal with highly non-linear and non-monotonic relationships between the response and the set of predictor variables. Like GLMs, the ability of this tool to handle non-linear data structures can aid in the development of ecological models that better represent the underlying data, and hence increase our understanding of ecological systems (Guisan *et al.*, 2002).

$$g(\quad_{(X)})\qquad_{0}\quad f_{1(X_1)}\quad\cdots\quad f_{m(X_m)} \tag{2.5}$$

Ahmadi-Nedushan (2006) reported a few applications of GAMs to aquatic habitat modelling described in the literature. Regarding macroinvertebrates in rivers, Milner *et al.* (2001) applied a GAM model to predict macroinvertebrate richness at the reach level across glacier-fed river sites.

Ridge regression and Principal component analysis: Ridge Regression (RR) and Principal Component Analysis (PCA) are particularly useful when the independent variables are highly correlated. In situations where multicollinearity exists, the challenge is to minimize the possibility of including redundant variables in the model. However, care must also be taken not to delete any important variables. One solution to the problem of multicollinearity is the use of the RR procedure. The use of RR is beneficial and preferable when the amount of data is not large relative to number of variables as noisy variables can frequently be included in the chosen subset (Ahmadi-Nedushan, 2006). This property makes RR a useful tool in habitat modelling as numbers of observations are often limited. PCA is another multivariate statistical method that can be used in case of multicolinear data. PCA is concerned with explaining the variance-covariance structure of a data set through a few linear combinations of the original variables. The purpose of PCA is to identify the dependence structure of multivariate observations in order to obtain a compact representation. The analysis identifies characteristic and uncorrelated modes of the variables. Ahmadi-Nedushan (2006) reported some applications of PCA used in aquatic habitat studies, mainly related with fish abundance but no one related with macroinvertebrates.

2.2.3.2. Artificial neural networks

Artificial Neural Networks (ANNs) are mathematical models based on the transfer of information through a network of functional units, called neurons. Given a number of input values, entered at the basis of the network, it generates one or more outputs. ANNs are non-linear mapping structures that can be applied for predictive modelling and classification. Various types of neural networks exist, suitable to solve different kinds of problems. The choice of the type of network depends on the nature of the problem to be solved. The most popular ANNs are multilayer feed-forward neural networks with the backpropagation algorithm (Goethals, 2005 and Goethals *et al.*, 2006). ANNs are currently recognized as an alternative for multivariate statistics to predict aquatic communities. An example of a relation can be the abundances of a number of macroinvertebrate taxa (such as Gammaridae, Tubificidae, Chironomidae) which are being predicted based on a number of environmental variables such as flow velocity, percentages of clay, silt and sand in the sediment, river depth, dissolved oxygen, pH among others. Based on such ANNs models, relations between individual variables (river characteristics) and probably of presence of aquatic macroinvertebrates can be determined (Goethals, 2005).

2.2.3.3 Classification trees

In contrast to ANNs, the application of classification trees in ecological modelling, in particular related to macroinvertebrates, is rather limited and hardly described in literature (Goethals, 2005). Classification trees, often referred to as decision trees, predict the value of a discrete dependent variable with a finite set of values (called class) from the values of a set of independent variables (called attributes), which may be either continuous or discrete. Data describing a real system, represented in the form of a table, can be used to learn or automatically construct a decision tree (Goethals *et al.*, 2006).

2.2.3.4 Fuzzy logic

To construct models for use in river management, mainly ecological monitoring data are used. However, such data often bear a large uncertainty, which is mostly not only epistemic uncertainty (e.g. measurement error, natural variation), but also includes linguistic uncertainty (e.g. vagueness). Sometimes the relations between the ecosystem

components are not exactly known and analytical models for establishing these relationships are not available or the data are insufficient for statistical analysis (Adriaenssens *et al.*, 2006). In such a case, a model can be build based on expert knowledge and a fuzzy logic approach used for solving uncertainty problems. Fuzzy systems use linguistic descriptions such as ‘low’, ‘high’ or ‘moderate’ to quantify variables and use ecological expert knowledge to transform these descriptions into a mathematical framework in which suitable data processing can be performed (Mouton *et al* 2009a). This turns fuzzy systems into a popular technique for ecological modelling, resulting in numerous applications. However, the main bottleneck in the application of fuzzy logic is the need for ecological expert knowledge (Mouton *et al* 2009a). Four models based on fuzzy logic were developed by Adriaenssens *et al.* (2006) for predicting macroinvertebrate taxa Asellus and Gammarus in the Zwalm river basin (Flanders, Belgium), based on an expert knowledge database and an ecological validation set with physical–chemical variables and macroinvertebrate monitoring data.

2.2.3.5 Bayesian belief networks

Bayesian Belief Networks -BBNs (Pearl, 1988) are probabilistic models that may be based on expert knowledge, empirical data, or a combination of both. BBNs are models with a network structure that focus on the explicit representation of ‘cause-and-effect’ relationships between variables, representing in this case ecosystem components. The network architecture is linked to probability distributions that allow it to deal with variability and uncertainty in the models. This is particularly useful for the description of ecological systems (Goethals, 2005). BBNs are probabilistic expert systems in which the knowledge base has two components: a causal network and a set of conditional probability matrices. The causal network defines the 'cause-effect' links between the variables. The conditional probability matrices define the probabilistic relationship that exists between the states of each ‘effect’ (or child) variable and the states of its ‘cause’ (or parent) variables. One of the first applications of this technique on macrobenthos communities was made by Adriaenssens *et al.* (2004). This paper describes a preliminary study evaluating the use of BBNs for prediction of two crustacean macroinvertebrate families, Asellidae and Gammaridae in rivers in Flanders, Belgium. Field data were used to represent the conditional probability relationships and expert judgement allowed the construction of the causal network.

2.2.4 Linking and integration of water system models

Robust ecosystem analysis of water resource systems remains elusive. A principle reason is the difficulty in linking engineering models used to simulate physicochemical processes associated with project design or operation with biological models used to simulate biological population attributes (Nestler *et al*., 2005). However, the coupling of ecological models to physical models is essential to make simulations on the effect of river alterations. For this, the set of available water system models (hydrological, hydraulic, water quality, etc.) limits the applications of the overall model set (Goethals, 2005).

Up to now, most ‘integrated’ ecological models are merely able to calculate water quality variables and some overall biomass figures about phytoplankton, macrophytes, fish communities, etc. without details about the composition of these communities. For this purpose, the use of habitat suitability models such as data driven or expert knowledge models, also sometimes referred to as mini-models can be useful for this more detailed type of calculations, where direct relations between a set of variables is calculated, without incorporating feedback loops (Goethals, 2005).

Some examples of ‘integrated’ ecological models for predicting phytoplankton, zooplankton, macrophytes or fish communities are: Nestler *et al.* (2005) with a coupled system illustrated by linking a fish swim path selection model with a two-dimensional hydrodynamic and water quality model (CE-QUAL-W2); Saito (1999) with a linked modelling approach of the model CE-QUAL-W2 linked with a bioenergetics model and a food web-energy transfer model, to investigate the effects of revised dam operations on the upstream reservoir ecosystem at Shasta Lake; Rodrigues *et al.* (2007) with an integrated modelling approach to simulate the three-dimensional dynamics of zooplankton, coupling a hydrodynamic model, the SELFE, and an ecological model based on an extension of EcoSim 2.0 model, which allows the simulation of several ecological state variables, to account for zooplankton dynamics; and Velez (2006) who presented an integration of water quality and ecosystems models using a two-dimensional hydrodynamic and water quality model (SOBEK Rural), to understand the complex relationships between the nutrient pollution and the growth of the water hyacinth and developed a framework to assess control strategies for this problem in the Sonso Lagoon in Colombia. The model was

developed in the Open Process Library of DELWAQ which facilitates the coupling online with the water quality model and the hydrodynamic model.

The application of ‘integrated’ ecological models for predicting macroinvertebrates in rivers, is rather limited and hardly described in literature. Mouton *et al*. (2009b) evaluated the strengths and weaknesses of the Water Framework Directive Explorer (WFD-Explorer) toolbox on the Zwalm River basin in Flanders, Belgium. This WFD-Explorer is a modular toolbox which supports integrated water management analysis in a river basin and it is mainly an appropriate tool for modelling the impact of different restoration measures on river ecology based on expert rules embedded in this simulation environment. The rules linked the physical and chemical water body characteristics to an Ecological Quality Ratio (EQR) for the different aquatic communities (fish, macrophytes and macroinvertebrates) of each water body.

3. MATERIALS AND METHODS

3.1 STUDY AREA

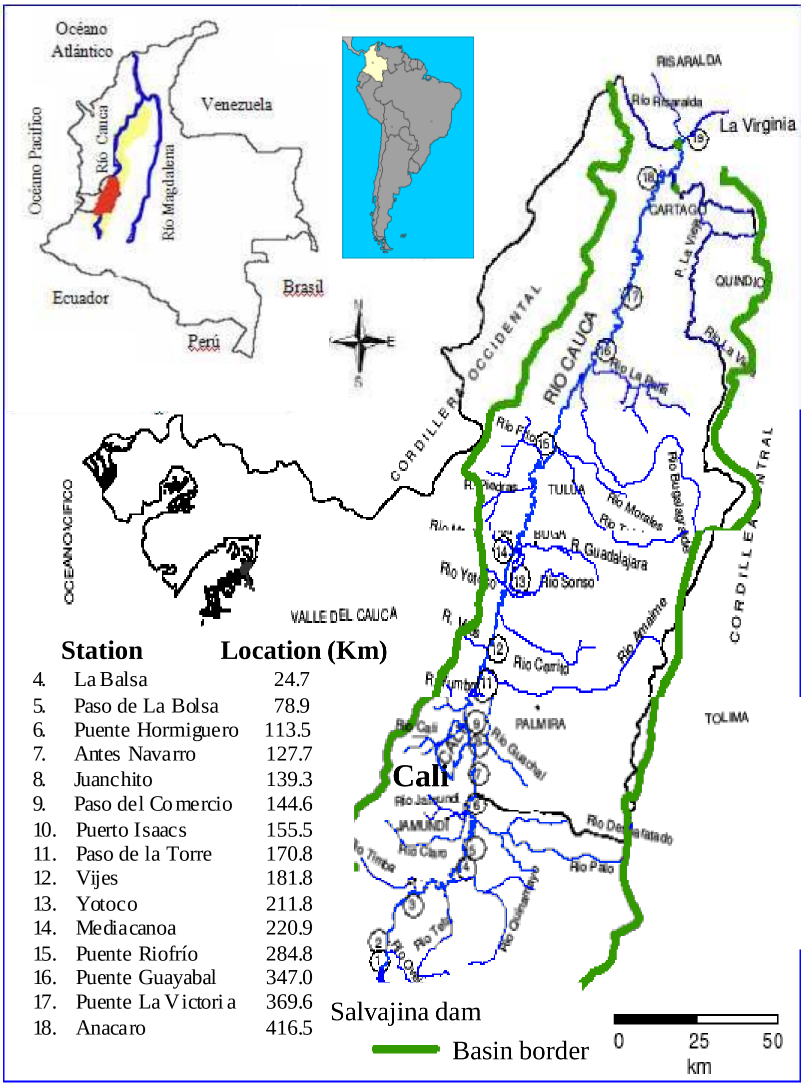
3.1.1 General aspects

The Cauca river is the second most important river in Colombia and the main hydric resource of the Colombian southwest. More than 16 million inhabitants live in its basin that corresponds to 38% of the Colombian population (CVC and Univalle, 2007a). The Cauca river geographical valley, located in the high basin, is especially important for the country's development and economy. A significant part of the south-western manufacturing industry, the paper production industry, the sugar cane agricultural industry and part of the coffee zone are located in this area. This study focuses on the stretch of the Cauca river which is situated in the geographical valley, which extends from the station La Balsa until Anacaro (Figure 3.1). The most important water quality problems in this river can be found in this zone, especially in the stretch close to the city of Cali.

3.1.2 Hydraulic, hydrological and morphological characteristics.

The Cauca river flows for 445 km in its geographical valley and descends from a height of 1000 meters (Salvagina dam) to 900 meters above sea level. This stretch of the river has an average width of 105 meters and it can fluctuate between 80 meters in the high part of its course (Salvajina dam– La Balsa) to 150 meters in the low part (Anacaro – La Virginia). The depth can vary between 3.5 and 8.0 m. The longitudinal profile of the river shows a concave shape with a hydraulic slope which oscillates between $7x10^{-4}$ m/m and $1.5x10^{-4}$ m/m (CVC and Univalle, 2007b). As it was said before, this study focuses on a stretch of the Cauca river located between the water quality monitoring stations La Balsa and Anacaro with a total length of 389 km.

In the Cauca river´s geographical valley there are mainly two climatic conditions with two periods each one, the rainy season (from March to May and from September to November) and the dry season (from December to February and from June to August), the last one with less precipitation. The average temperature is 24ºC, with fluctuations in the range from 10ºC to 38ºC. The mean annual precipitation in the study zone is approximately 1000 mm. (CVC and Univalle, 2007a).



Source: CVC and Univalle, 2007b

Figure 3.1 Drainage basin of the Cauca river´s geographical valley

3.1.3 Environmental aspects

The main environmental authority in the Cauca river´s geographical valley is the Corporación Autónoma Regional del Valle del Cauca (CVC), which has been working in the water quality monitoring program of the Cauca river and its tributaries since the 1970s. The main physical-chemical and bacteriological parameters are monitored four (4) times per year. These data, enriched along the years, have supported the first simulations of quality of the water. In the year 1985 the Salvajina dam entered in operation, located at the beginning of the geographical valley, with the main aim of flood control. Furthermore, hydro-electrical energy generation and pollution control during the dry season by the increase of the minimum river flow (higher dilution effect of wastewater discharges) were taken as secondary aim. In general, flow values and the water quality physical-chemical

parameters of the Cauca river in the study zone, are influenced by the hourly fluctuation of the domestic and industrial wastewater discharges into the river and the operation of the Salvajina dam. In the last case the hydro-electrical energy generation causes hourly fluctuations of the river flow and therefore hourly changes in the water quality due to the increase or decrease of the dilution capacity of the river, mainly in the stretch La Balsa – Juanchito (CVC and Univalle, 2007a).

According to CVC and Univalle (2007a) and CVC (2005) the Cauca river during the year 2005 received in average 195 tons of organic matter load per day in terms of BOD_5 in the study zone. As can be seen in the Figure 3.2, the main source of water pollution in the study zone is related with the discharge of domestic wastewaters, with around of 72.6% from the total of organic matter load discharged into the river in the Valle del Cauca department. The city of Cali contributed with 75 tons of BOD_5 per day, and the rest of the municipalities of the department discharged into the river in average 66.76 tons of BOD_5 per day. Only 5 of the 33 municipalities of the Valle del Cauca department, which are located in the Cauca's river basin, had wastewater treatment systems. Regarding to the industrial wastewater discharges, they are characterized by their content of organic and inorganic pollutants. Some of these industries are metal processing, tanneries and gold mining which generate many toxic substances such as trivalent chromium, sulphurs, heavy metals, mercury and cyanide, which are highly toxic for humans and aquatic biota CVC and Univalle (2007a). In Annex A the location of the domestic and industrial wastewater discharges, which are monitored by the CVC in the Cauca´s river basin, is presented.

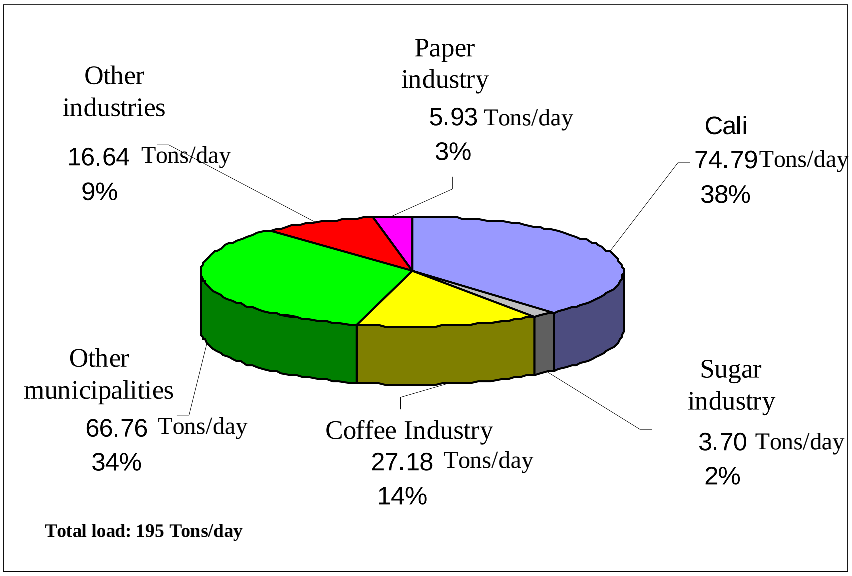


Figure 3.2 Organic matter load in terms of BOD_5 discharged into the Cauca river in the stretch La Balsa – Anacaro. Year 2005. Source: CVC (2005)

3.2 DATA AND INFORMATION COLLECTION TO DEVELOP AND VALIDATE MODELS

The database used in this research corresponds to the information collected and analyzed during the years 1997-2007 by the CVC and the Cauca River Modeling Project (PMC). The information related to benthic macroinvertebrates and physical-chemical parameters in the study zone in the Cauca river, was presented by CVC-Univalle (2004c) as well as the information related to the calibration, implementation and application of the hydrodynamic and water quality model of the Cauca river (model MIKE11) (CVC-Univalle, 2007a and 2007b).

The database used for the implementation of the ecological model for predicting macroinvertebrates in the Cauca river was selected considering the simultaneousness of biological, physical-chemical and hydraulic measurements in the sample stations. Therefore from the information presented by CVC-Univalle (2004c) (Table 3.1) only the database from the years 1997 (Paso de la Balsa, Paso de la Bolsa, Juanchito, Paso de la Torre, Mediacanoa during February), 2001 (Anacaro) and 2004 (Paso de la Balsa, Paso de la Bolsa, Puente Hormiguero, Antes Navarro, Juanchito, Paso de la Torre, Mediacanoa, Puente La Victoria and Anacaro) was used. In Annex B the database selected for the implementation of the ecological model for predicting macroinvertebrates in the Cauca river is presented. The simultaneousness of the biological, physical-chemical and hydraulic measurements in the sample stations is a critical issue because this database feature allows establishing reliable statistical relationships for the ecological model.

The procedure for collecting benthic macroinvertebrates in the Cauca river was developed considering the type of substrate: stone substrate, sand substrate, mud substrate, trunks and floating vegetation substrate (Figure 3.3). Thus all type of macroinvertebrate habitats were considered during the monitoring. Macroinvertebrates in the stone substrate were found by hand net sampling method mainly in the station La Balsa, located after the Salvajina dam where high water flow velocities are present. For the sand substrate a grab sampling method was carried out, and samples were collected in the middle and at the borders of the transversal section. For the mud substrate two methods were performed, a grab sampling method and a hand net sampling method. For the trunks and floating vegetation substrate a manual sampling method was carried out. After the samples were collected in each zone

and substrate in the Cauca river, they were preserved, labeled and transported to CVC's environmental laboratory where finally an identification and classification of macroinvertebrates, mainly at the taxonomic level of genus (phylum, class, order, family and genus) was carried out (CVC-Univalle, 2004c).

Table 3.1 Biological, physical-chemical and hydraulic information in the sample stations in the Cauca river.

Year	Month	Sample Stations								
		Paso de La Balsa	Paso de La Bolsa	Puente Hormiguero	Antes Navarro	Juanchito	Paso de La Torre	Mediacanoa	Puente La Victoria	Anacaro
1996	November	X	X			X				
	December	*	*			*				
1997	February	X / *	X / *			X / *	X / *	X / *	*	*
	March									
	April								X	
	May									X
1998	February						*	*		
	March						X	X		
	August					X				
	November					*				
1999	November									X
	December									*
2000	February	X	X			X				
	Mrach						X	X	X	X
	April	*	*			*		*	*	
	May									
	June									
	July	*	*			*	*	*	*	*
2001	June									X / *
2004	June	X / *	X / *	X / *	X / *	X / *	X / *	X / *	X / *	X / *

X : Biological information
* : Physicochemical information
Hydraulic information is available in all sample stations by measurements or interpolations with the hydrodinamic model

Source: CVC-Univalle, 2004c

Simultaneously with the biological monitoring, water samples were taken for the analysis of physical-chemical parameters. These water samples were preserved at 4°C and transported to the CVC's environmental laboratory for the analysis (Figure 3.4). The hydraulic information reported by the sample stations during the monitoring campaigns was mainly provided by CVC from hydrometric stations. When this information was not available interpolations with the hydrodynamic model (MIKE11) were performed.

Applications of the integrated hydrodynamic, physical-chemical and ecological water quality model, developed for the Cauca river for predicting macroinvertebrate communities present in this river under different conditions, were carried out considering scenarios for the pollution control (see CVC and Univalle, 2007a) proposed by the environmental

authority (CVC), the municipalities and the industries in the Cauca river´s geographical valley.

Figure 3.3 Procedure for collecting benthic macroinvertebrates in the Cauca river considering the type of substrate. A. Grab sampling method for sand substrate and mud substrate; B and C. Manual sampling method for trunk and floating vegetation substrate; D and E. Hand net sampling for mud substrate. Source: CVC-Univalle (2004c).

Figure 3.4. Procedure for collecting water samples for the analysis of physical-chemical parameters (F) and identification and classification of macroinvertebrates (G and H). Source: CVC-Univalle (2004c).

3.3 WATER QUALITY MODELLING SOFTWARE USED IN THE CATCHMENT OF THE CAUCA RIVER

3.3.1 Hydrodynamic and physical-chemical water quality models

The Cauca river physical-chemical water quality and hydrodynamic model (MIKE11) used in this research is a mathematical simulation model which was calibrated and verified for dynamic flow conditions in the framework of the Cauca River Modelling Project (PMC) (CVC-Univalle, 2007a and 2007b). This model reproduces in acceptable form the values of DO, BOD_5 and temperature in the monitoring stations of the Cauca river, considering hourly fluctuations. The model is conformed by 62 cross sections, 2 external boundaries (La Balsa and La Virginia), 96 internal boundaries which include 27 rivers and streams, 9 municipal wastewater discharges, 12 industrial wastewater discharges and 37 water extraction sites. Each internal boundary was represented like a lateral extraction or discharge. Water quality modelling was carried out in Level 1 of MIKE11, which includes temperature, BOD, and DO as state variables. In the framework of the PMC project two monitoring campaigns with calibration and verification purposes for the water quality model were carried out during the months of August of 2003 and February of 2005. These campaigns had a duration of respectively five (5) and four (4) 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, COD, TSS).

3.3.2 Ecological model for predicting macroinvertebrates in the Cauca river

In general, two approaches can be followed when performing (ecological) modelling which are mechanistic and data driven modelling. Which approach will be preferable in a specific case depends on the aim of the research, the knowledge of the ecological processes and state variables in the system, the required properties of the model and the data set available. In ideal conditions with complete and reliable information, the application of mechanistic models for predicting biological species in aquatic ecosystems is the best choice. However, in real conditions researchers have sometimes to deal with lack of information related with the data set available, the ecological processes and state variables in the system. Furthermore, mechanistic models are merely capable to calculate water quality variables and some overall biomass information about the biota present. Details on

the composition of these communities are mostly modelled by data driven models (De Laender *et al.*, 2005 and Goethals, 2005).

In the case of the Cauca river the water quality model MIKE11 allows to calculate water quality variables such as temperature, BOD, and DO. However, there is a lack of information about processes associated with particulate organic matter and nutrients, which are essential in case of using ecological mechanistic models (i.e. food-webs). Considering these limitations, the use of habitat suitability models, such as data driven (e.g. statistical models) for predicting macroinvertebrates in the Cauca river, can be useful. Additionally, these models can be used as a first approach for modelling composition of macroinvertebrate communities, and they can be useful for this more detailed type of calculations, where direct relations between a set of predictor variables (physical-chemical and hydraulic) and biological species are calculated, without incorporating feedback loops.

The choice of the appropriate model of species-environment relationship depends on the goals and resources of the study and especially on the types of measured environmental and response variables. The use of classification trees, Artificial Neural Networks (ANNs), fuzzy logic, Bayesian Belief Networks (BBNs) and Generalized Additive Models (GAMs); is preferable when the species-habitat relationships are nonlinear, which is the case for most commonly used data. However, a drawback of these methods, all non-parametrical methods, is that they do not provide users with a conventional mathematical function and it is difficult to export a model. This problem becomes particularly evident when the aim is to build a spatial prediction in a Geographic Information Systems (GIS).

Other methods, such as multiple linear regression and Generalized Linear Models (GLMs) (mathematical extensions of linear models for non-linearity and non-constant variance structures in the data), both parametrical methods, provide users with a conventional mathematical function, which allow them to build models for predicting the probability of species occurrence or estimating numbers of organisms (abundance) at new sites and/or future times. However, where the assumptions of normality inherent in linear regression cannot be justified, modern regression modelling paradigms like GLM are more appropriate. Thus, GLMs are more flexible and better suited for analyzing ecological relationships, which can be poorly represented by classical Gaussian distributions, therefore GLMs can be used as effective tools for the analysis of aquatic habitat-species

relationship, especially when the assumptions of simpler methods cannot be justified (Guisan *et al.,* 2002). Considering the recommendations described in literature for ecological modelling and the aim of this research, for the integration of the hydrodynamic and physical-chemical water quality model MIKE11 with an ecological model for predicting macroinvertebrates in the Cauca river, the parametric approach using GLMs is the best choice.

In GLMs the predictor variables $X = (X_1,\ldots,X_m)$ are combined to produce a linear predictor LP which is related to the expected value $E(Y)$ of the response variable Y through a link function g (), such as:

$$g(E(Y)) \quad LP \quad {}_0 \quad X^T \qquad (3.1)$$

where $E(Y)$ is the expectation of Y; ${}_0$ is a constant called the intercept, $X = (X_1,\ldots,X_m)$ is a vector of m predictor variables and ${}_1,\ldots {}_m$ is the vector of m regression coefficients (one for each predictor). Thus, the model for generic variables X and Y, considering the corresponding terms for the *ith* observation in the sample is:

$$g({}_i) \quad {}_0 \quad {}_1X_{i1} \quad {}_2X_{i2} \quad \ldots \quad {}_mX_{im} \qquad (3.2)$$

Unlike classical linear models, which presuppose a Gaussian (i.e. normal) distribution and an identity link, the distribution of the response variable (*Y)* in a GLMs may be any of the exponential family distributions (e.g. Gaussian, Poisson or binomial) and the link function may be any monotonic differentiable function (like logarithm or logit). In GLMs the unknown parameters, β, are typically estimated with maximum likelihood, maximum quasi-likelihood, or Bayesian techniques. These techniques are commonly available in software packages such as S-Plus, SAS or R.

Of the GLM techniques, Logistic regression (i.e. GLM with logit link function and binomial error distribution) is the most frequently used modelling approach for predicting the probability of species occurrence (Manel *et al.,* 2000; Pearce and Ferrier, 2000; Ahmadi-Nedushan, 2006; Rushton *et al*., 2004; Reineking & Schroderc, 2006), because a single record of presence or absence of the target species can be considered to be a binomial trial with a sample size of 1. This method has become a favourite tool in habitat

modelling when the species information is given as presence/absence data, because this information is comparatively easy to collect in the field, even when the zero data set has to be created afterwards by a different sampling strategy.

Logistic regression can be used to analyse the relationship between a Bernoulli (or binary) response (values restricted to 1 or 0, i.e. suitable (1) versus unsuitable (0)) and explanatory environmental factors describing the quality of the habitat (e.g. depth, velocity and DO). Logistic regression allows for the simultaneous analysis of categorical (e.g. substrate, cover) and continuous variables such as depth and current velocity (Ahmadi-Nedushan, 2006). The model estimates the probability of a positive response occurring given a set of explanatory variables. 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, 2006). Logistic regression model is expressed as:

$$g_{(X)} = Logit(p) = \ln\left[\frac{p_i}{1-p_i}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m \qquad (3.3)$$

$$p = \frac{e^{g_{(x)}}}{1+e^{g_{(x)}}} = \frac{1}{1+e^{-g_{(x)}}} = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\dots+\beta_m X_m)}} \qquad (3.4)$$

where *g(x)* is the linear combination of environmental factors; p_i is the probability of species presence in a cell or the probability that a habitat cell would be suitable for species; β_0 and $\beta_1, \ldots, \beta_m$ are regression constants, and $X = (X_1, \ldots, X_m)$ is a vector of *m* predictor variables. Transforming the probability to odds removes the upper bounds, and taking the logarithm of the odds removes the lower bound.

In order to select the best logistic regression model (combination of predictor variables) for predicting the probability of macroinvertebrate species occurrence in the Cauca river the Akaike's information criterion (AIC) was used. Models were fitted using the maximum likelihood method (McCullagh & Nelder, 1989) with backwards elimination to select the final predictor variables. The step function, implemented in the statistical software XLSTAT version 2009 (Addinsoft, 2009) used in this research provides a procedure for

this purpose using the AIC; this is a penalized version of the likelihood function in which the best model is given by the lowest AIC value. The use of AIC in GLM is associated with identifying variables for inclusion or exclusion in models. AIC is actually equivalent to twice the log-likelihood of the model fitted plus two times the number of parameters estimated in its formation (Rushton *et al*., 2004). Given that the model with the smallest log-likelihood is considered to be that with the best fit, the addition of two times the number of parameters means that AIC effectively includes a penalty for adding predictor variables to the model. Thus, AIC aids identifying the most parsimonious model amongst a set. These approaches have been used in vertebrate studies particularly. Manel *et al*. (1999); Williams & Ormerod (2001) used it for birds and Jaberg & Guisan (2001), Schadt *et al*. (2002) and Johnson, Seip & Boyce (2004) for mammals (Rushton *et al*., 2004).

For estimating numbers of organisms (modelling abundance) at new sites and/or future times, other modelling approaches based on GLM techniques, such as Poisson regression, quasi-Poisson regression and negative binomial regression (Jones *et al.*, 2002; Ver Hoef, and Boveng, 2007; Guisan *et al.*, 2002; White and Bennetts, 1996; Dolan *et al.*, 2000; Sudhir and Krishna, 2007) are frequently used.

Often, data on organisms come in the form of counts, and in ecological modelling the idea is to relate these counts to environmental conditions. Linear regression is commonly used, but this approach is not the most appropriate for count data, which are nonnegative integers (i.e. 1, 2, 3, ...). Therefore, other kind of regression models based on Poisson or negative binomial distributions are commonly used. Count data in ecology are often ‘‘overdispersed’’ (i.e. for any data set or model the variance exceeds the mean). For a Poisson distribution, the variance is equal to the mean. This may be quite restrictive for biological data, which often exhibit more variation than given by the mean. A common way to deal with overdispersion for counts is to use a GLM framework (McCullagh and Nelder 1989), where the most common approach is a ‘‘quasi-likelihood,’’ with Poisson-like assumptions (hereafter simply called quasi-Poisson) or a negative binomial model, because they are widely available in software and they generalize easily to the regression case (Ver Hoef, and Boveng, 2007). These approaches depend on a distributional form and likelihood; however, quasi models are only characterized by their mean and variance and allow to estimate the parameters in a model without specifying the underlying distribution function. The quasi-likelihood methods has the advantage of relative computational

simplicity, speed and robustness, as they can make use of the more straightforward algorithms developed to fit GLMs, and additionally they have advantage of leaving parameters in a natural, interpretable state and allows standard model diagnostics without a loss of efficient fitting algorithms (Ver Hoef, and Boveng, 2007).

In quasi-Poisson regression models (QPRM) the mean, μ, of the distribution depends on the independent variables, X, through:

$$E(Y) = \mu = g^{-1}(X\beta) \qquad (3.5)$$

$$Var(Y) = \theta V(\mu) = \theta V_{qPoi}(\mu) \qquad (3.6)$$

where $E(Y)$ is the expectation of Y, also known as the ''mean'' of the distribution; g^{-1} is the link function; X is a design matrix of both continuous and categorical covariates; β is a vector of parameters (regression coefficients); θ is a scale parameter (also known as overdispersion parameter); and $Var(Y)$ is the variance of Y, with μ>0 and θ>1. Although μ>0, the data themselves can be any nonnegative integer. The close relationship between Eq. 3.6 and the expectation and variance of a Poisson distribution, along with the use of a log link function, justifies calling this a ''quasi-Poisson'' model, denoted as $Y \sim qPoi(\mu, \theta)$ (Ver Hoef, and Boveng, 2007). When the scale parameter is expected to be higher than the value anticipated under the chosen distribution (i.e. overdispersion), the scale parameter can be estimated using quasi-likelihood; an extension of generalized least-squares (Guisan *et al.*, 2002).

For quasi-Poisson regression, it is assumed $Y_i \sim qPoi(\mu_i, \theta)$, where the mean μ_i for the *ith* observation vary as a function of the covariates for that observation. Because the mean μ_i >0, it is natural to model quasi-Poisson regression as:

$$\mu_i = e^{\beta_0 + \beta_1 X_{1,i} + \cdots + \beta_m X_{m,i}} \qquad (3.7)$$

The unknown parameters, β, in the QPRM for modelling macroinvertebrate species abundance in the Cauca river were estimated by means of the maximum quasi-likelihood method using the statistical software S-PLUS version 6.1. In order to select the explanatory

variables in the QPRM, changes in goodness of fit statistics were used to evaluate the contribution of subsets of explanatory variables to a particular model. The deviance (i.e. how much variation is left), defined to be twice the difference between the maximum attainable log likelihood and the log likelihood of the model under consideration, was used as a measure of goodness of fit. The strategy followed for variable selection was to fit a sequence of models, beginning with a simple model with only an intercept term, and then include one additional explanatory variable in each successive model. The importance of the additional explanatory variable was measured by the difference in deviances or fitted log likelihoods between successive models. More details about this technique can be found in Gordon, 2008.

4. RESULTS

4.1 RIVER WATER QUALITY ASSESSMENT

4.1.1 Physical-chemical parameters

The Cauca river is one of most severe cases of contamination for domestic and industrial wastewater discharges in Colombia. According to CVC (2005), the Cauca river received during the year 2005 in average 195 tons of organic matter load per day in terms of BOD_5 in the study zone. The rapid urbanization and major economic development in the Cauca river´s geographical valley has led to dramatic degradation of the environment and increased health risks due to inefficient processing of the increased pollutant load effluents and solid wastes. A major example of this is the city of Cali, with a population of more than two million inhabitants and limitations of the treatment of its wastewaters. This city discharged during the year 2005 around 75 tons of BOD5 per day into the Cauca river. The high discharge of organic matter causes a high oxygen demand in the degradation process. Thus, one of the most sensitive problems in the Cauca river is the decrease of dissolved oxygen (DO). The stretch Hormiguero (km 113.5) – Mediacanoa (km 220.5) is the zone that receives the largest polluted discharges from the city of Cali (Photo 4.1) and the industrial zones of Yumbo and Palmira cities. This pollution problem is critical after the river crosses the city of Cali, especially during dry season (low flows), when pollution can reach values of 7.6 mg/l of BDO_5 and concentrations of DO near to zero (0) mg/l (Figure 4.1).

Photo 4.1 Impact of domestic wastewater discharges generated in the city of Cali in the Cauca river. Source: CVC and Univalle (2004d).

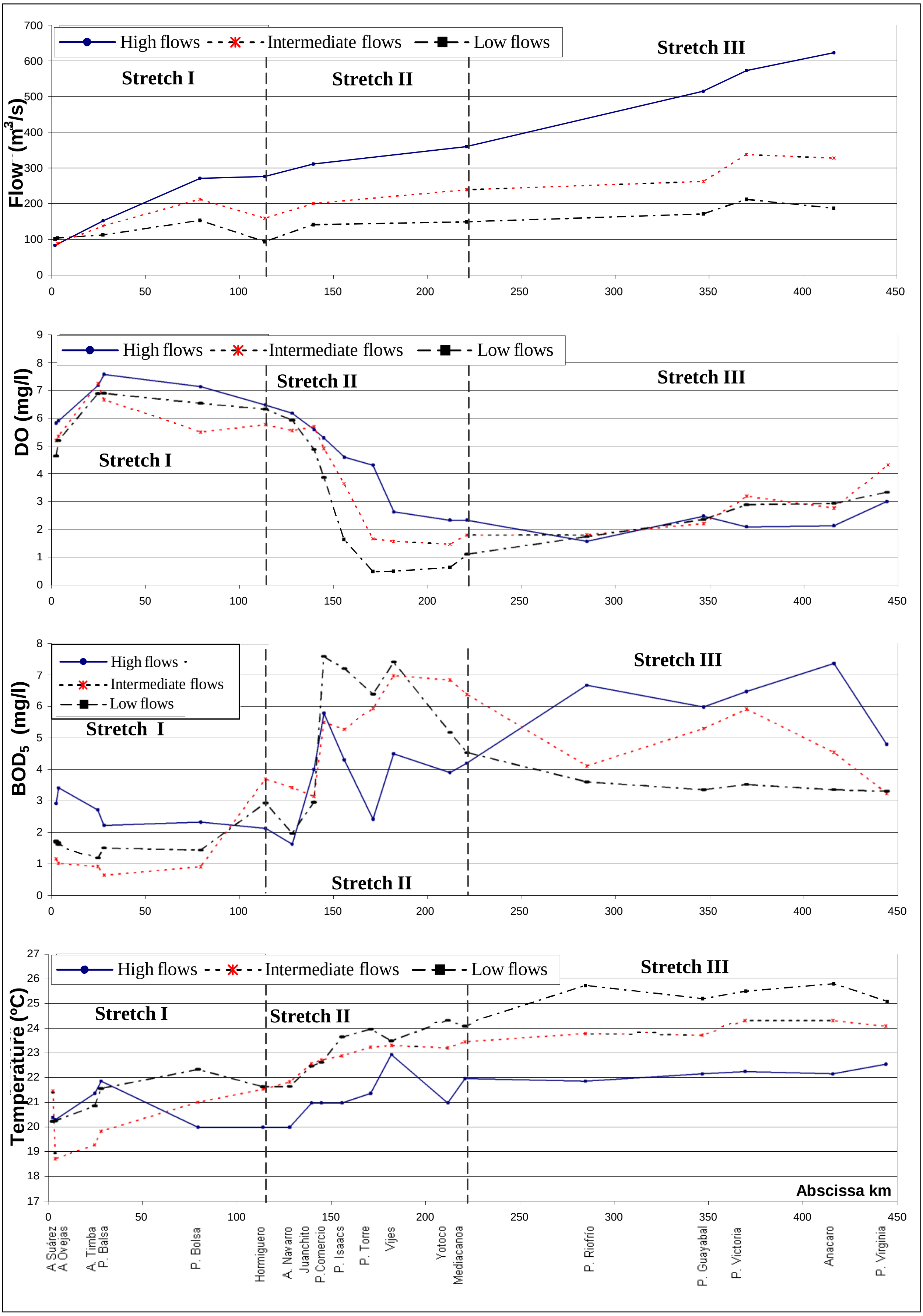


Figure 4.1 Variation of flow, DO and BOD_5 in the Cauca river. Period 2001-2003. High flows, intermediate flows and low flows conditions. Stretch Salvajina dam – La Virginia. Source: CVC and Univalle (2004d).

In order to analyze the water quality in the Cauca river it was necessary first to assess the effect of the climatic conditions on it. In the framework of the PMC project the historical water quality database was classified in 3 groups (high, intermediate and low flows) considering the river flow associated to the date of the monitoring campaigns. The high flow condition was defined as the state in which the flow presented permanence lower than the 30 % of the time (i.e. flow ≥ 292 m^3/s), according to the duration flow curve in a specific station (e.g. considering as reference the river flow at Juanchito in the Figure 4.2). For the low flow condition it was considered a permanence time higher than 70 % of the time (i.e. flow ≤ 179 m^3/s), whereas for the intermediate flow condition the range between the 30% and the 70% (i.e. flow between 179-292 m^3/s) of permanence was adopted.

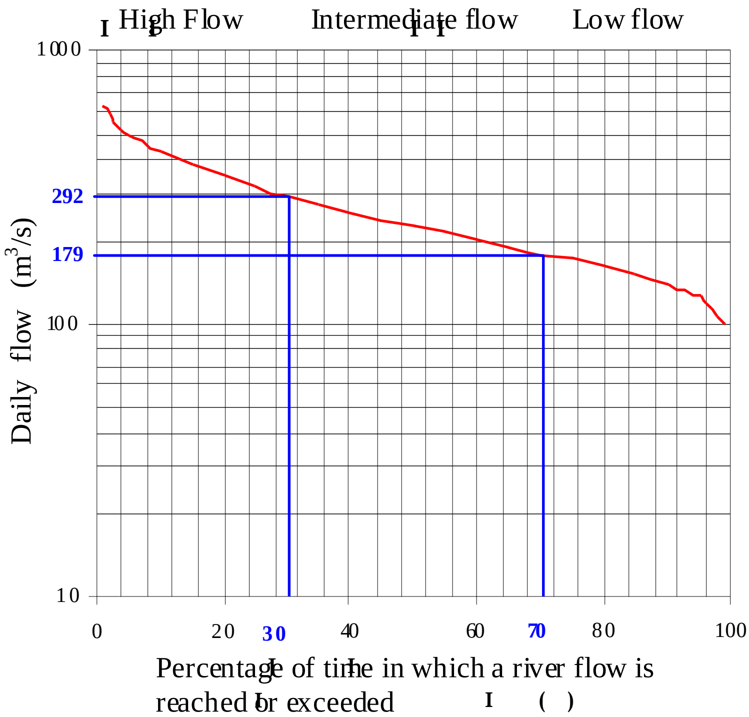


Figure 4.2 Cumulative frequency, or flow-duration, curve for the Cauca´s river at Juanchito station. Source: CVC and Univalle (2007c).

Taking into consideration the water quality of the Cauca river, the study zone can be divided in three (3) stretches (Figure 4.1). The first stretch located between the stations La Balsa and Hormiguero, where the river presents a good water quality with DO concentrations above 5.2 mg/l (70 % DO saturation concentration, for the Cauca river) and BOD_5 concentrations below 2.2 mg/l, in the three ranges of flows (high, intermediate and low flows). The second stretch located between the stations Hormiguero and Mediacanoa, is the most critical zone in terms of pollution, especially during low flow conditions, where the DO values are lower than 1 mg/l between the stations Paso de La Torre and

Mediacanoa, and the BOD_5 concentration reaches a maximum value of 7.6 mg/l. The water quality deterioration in this stretch is related with the high organic matter loads in terms of BOD_5, coming from domestic and industrial wastewater discharges into the river from the cities of Cali, Yumbo and Palmira. According to Chapman (1996), concentrations of DO 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. Finally, the third stretch located between the stations Mediacanoa and Anacaro, where the river shows a slight recuperation of the water quality, which could be related with the degradation of the organic matter (self-cleaning capacity) and an increase of the dilution effect generated by the discharge of tributary rivers in this zone. In spite of the increment observed in the DO concentrations in this stretch in the three ranges of flows, their values are still lower than the minimum standard value established by the Colombian Decree 1594/84 for different uses of the water resource, which means, smaller than 70% of the DO saturation concentration (5.2 mg/l for the Cauca river). The water temperature in the 3 stretches of the studied river fluctuated between the normal range (0-30ºC) reported by the literature (Chapman, 1996). The water temperature profile increases downwards due to the decrease of the altitude throughout the Cauca's geographical valley.

4.1.2 Physical-chemical water quality index – ICAUCA index

Taking into consideration that many of the Water Quality Indices (WQIs) used worldwide have been developed considering specific water quality criteria and characteristics present in the countries or regions where they were developed, in the framework of the PMC project an adaptation of a WQI specific for the Cauca river, called ICAUCA, was developed based on some of the most frequently WQIs used worldwide (i.e. ICA-NSF and ICA-Dinius) and those used in developing countries in Latin America (i.e. ICA-CETESB for rivers in Sao Pablo Brazil and the ICA-Rojas for the Cauca river in Colombia) and its own environmental conditions (CVC and Univalle, 2004b).

The physical-chemical and microbiological water quality parameters included in the ICAUCA calculation are: DO, BOD_5, suspended solids, total solids, turbidity, colour, pH, total phosphorous, total nitrogen and faecal coliforms. The importance of these parameters in the ICAUCA calculation can be seen in the Table 4.1. The ICAUCA reduces this great

amount of parameters into a simple mathematical expression which represents all evaluated parameters and it enables easier interpretation of monitoring data. The final result of the ICAUCA is a value between zero (0) and one hundred (100), which can be expressed as well as a verbal description (e.g. a value of one hundred (100) for very good quality and a value of zero (0) for very bad water quality), a symbol or a colour (e.g. blue for optimum quality and red for very bad water quality). The water quality classification ranges according to the ICAUCA for drinking water use previous treatment can be seen in the Table 4.2.

Table 4.1 Weighing of the parameters included in the ICAUCA calculation

Parameter	Weighing (%)
Dissolved oxygen	21
BOD_5	15
Suspended solids	5
Total solids	7
Turbidity	7
Colour	5
pH	8
Total phosphorous	8
Total nitrogen	8
Faecal coliforms	16

Source: CVC and Univalle (2004b).

Table 4.2 Water quality classification ranges according to the ICAUCA for drinking water use previous treatment

ICAUCA value	Water quality classification	Classification colour
80-100	Optimum	Blue
50-80	Good	Green
35-50	Acceptable	Yellow
20-35	Unsuitable	Orange
0-20	Extremely bad	Red

Source: CVC and Univalle (2004b).

The complete results of the water quality assessment of the Cauca river using the ICAUCA can be seen in the Annex C. The Figure 4.3 shows the ICAUCA profile throughout the river, considering the most complete physical-chemical, biological and hydraulic database for a same year (i.e. monitoring campaign carried out at June of 2004, which corresponds to low flow conditions). In general, the ICAUCA values show that the water quality in the

Cauca river never reaches an optimum level for drinking water use previous treatment (i.e. ICAUCA > 80) at the monitoring stations at the 3 stretches. Additionally, the ICAUCA shows progressive water quality deterioration, starting with a good quality level at the first stretch, which progressively reaches an extremely bad water quality level at the Paso de La Torre station, in the second stretch, which is the most critical zone in terms of pollution. Finally, the ICAUCA shows a slight recuperation of the water quality at the third stretch, reaching an acceptable water quality level. The water quality assessment using the ICAUCA confirms the evaluation made previously with individual physical-chemical parameters (DO and BOD_5).

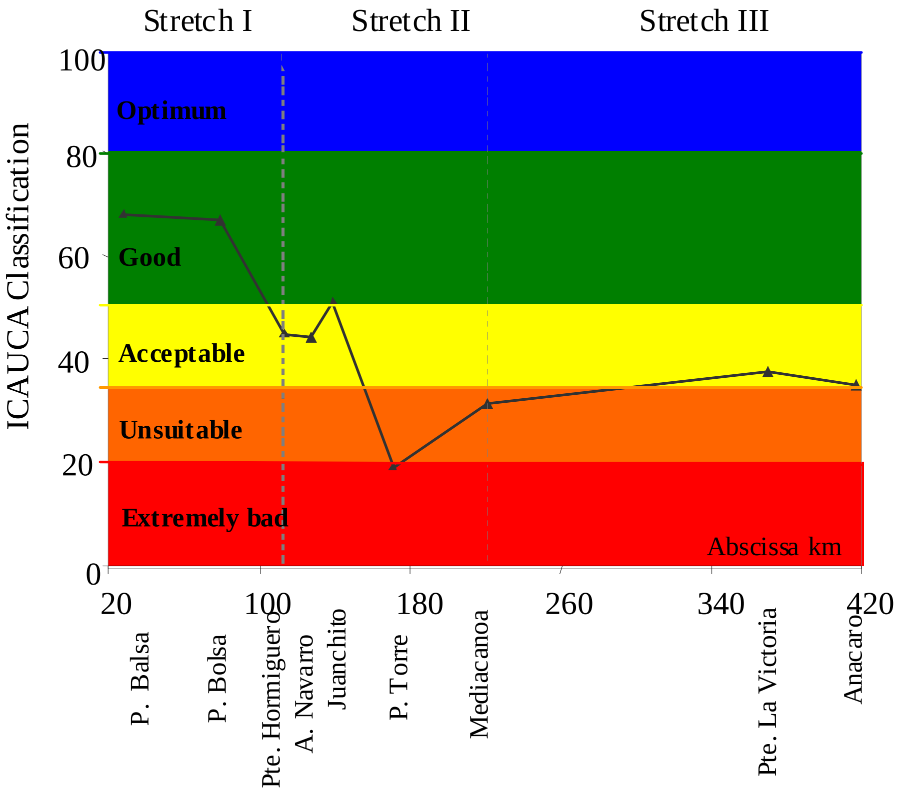


Figure 4.3 Water quality assessment of the Cauca river using the ICAUCA index considering. Monitoring campaign year 2004.
Source: CVC and Univalle (2004b).

4.1.3 Ecological indices and their relation with the ICAUCA index

The inputs of sewage and industrial effluents generate the establishment of a longitudinal succession of physical conditions and biota downstream of the effluent outfall. This can be seen quite clearly for organic pollutants arising from sewage effluent discharge where the effluent makes up a considerable proportion of the stream volume. The diagrammatic representation of these downstream changes can be seen in the Figure 4.4. This pattern is

typical for bigger rivers while in small streams the pattern is more variable. The benthic fauna can be more heavily impacted closer to the discharge point than the oxygen minimum due to deposits of organic matter or bacterial growths on the riverbed.

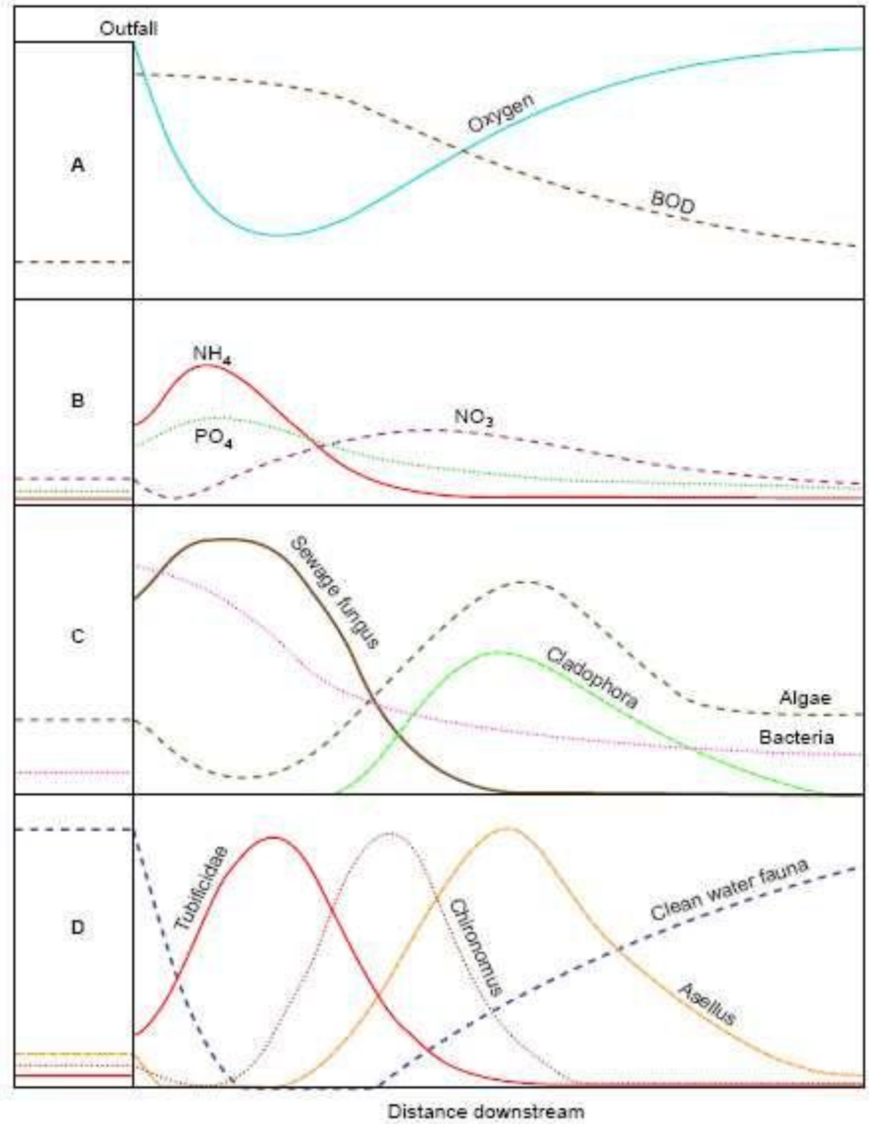


Figure 4.4 A diagrammatic representation of the longitudinal zonation established downstream the outfall of a continuous organic effluent discharge: (A) and (B) are physical and chemical changes; (C) changes in microorganism and plants; (D) changes in larger organisms (macroinvertebrates).
Source: modified from Hynes (1960).

In response to the physical-chemical changes (DO, BOD_5, NH_4, NO_3 and PO_4 variation) generated by the effluent outfall, a longitudinal zonation or succession of microbes, algae, and animals (macroinvertebrates) is produced, with the most pollution-tolerant species near the outfall (e.g. Tubificidae) and gradual recovery of the pollution-sensitive, clean water, fauna further downstream as the physical-chemical environment recovers (Giller and

Malmqvist, 1998). These relationships between pollution levels and the river community have been used in the development of a range of biotic indices, particularly successful for organic pollution, which are used for monitoring water pollution. The complete results of the water quality assessment of the Cauca river using ecological indices can be seen in the Annex C. In the following paragraphs a discussion of the results of ecological indices estimated, considering the most complete physical-chemical, biological and hydraulic database for a same year (i.e. monitoring campaign of 2004) will be developed.

4.1.3.1 Abundance of species (*N*)

This ecological index, represented by the letter *N*, is equivalent to the total number of organisms collected in each monitoring station. In the Figure 4.5 can be seen that whereas the water quality decreases, the amount of pollution-tolerant species (e.g. Tubificidae) increase and vice versa. The highest abundance and the lowest ICAUCA value were presented in the second stretch, in the Paso de La Torre station, confirming that this is the most critical zone in terms of pollution.

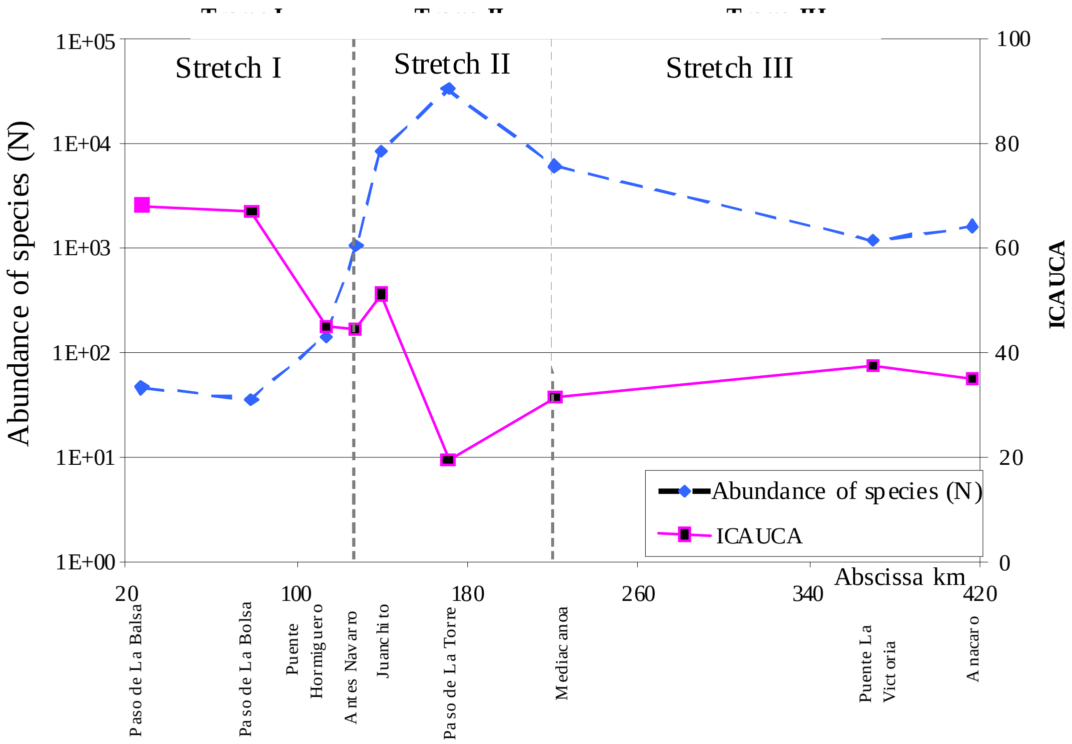


Figure 4.5 Variation of the abundance of species and the ICAUCA index throughout the Cauca river, during the monitoring campaign carried out during June of 2004. Source: CVC and Univalle (2004c).

In response to the aquatic ecosystem changes, such as those generated by industrial and domestic wastewater discharges, as happens in the Cauca river, the environmental conditions are disturbed and for many species the alternative is “to survive” or “to die”. The final result is a simplification of the communities, with an increase of individuals in the populations which had the capacity “to adapt” or “to survive” to the new conditions (Roldan, 1992).

4.1.3.2 Richness of species (*S*)

The richness of species (*S*) at the Cauca river monitoring stations was estimated by means of direct counting of the number of species in each site. From the Figure 4.6 can be seen that the lowest *S* values were presented at the second stretch of the Cauca river, where the lowest ICAUCA values were reached.

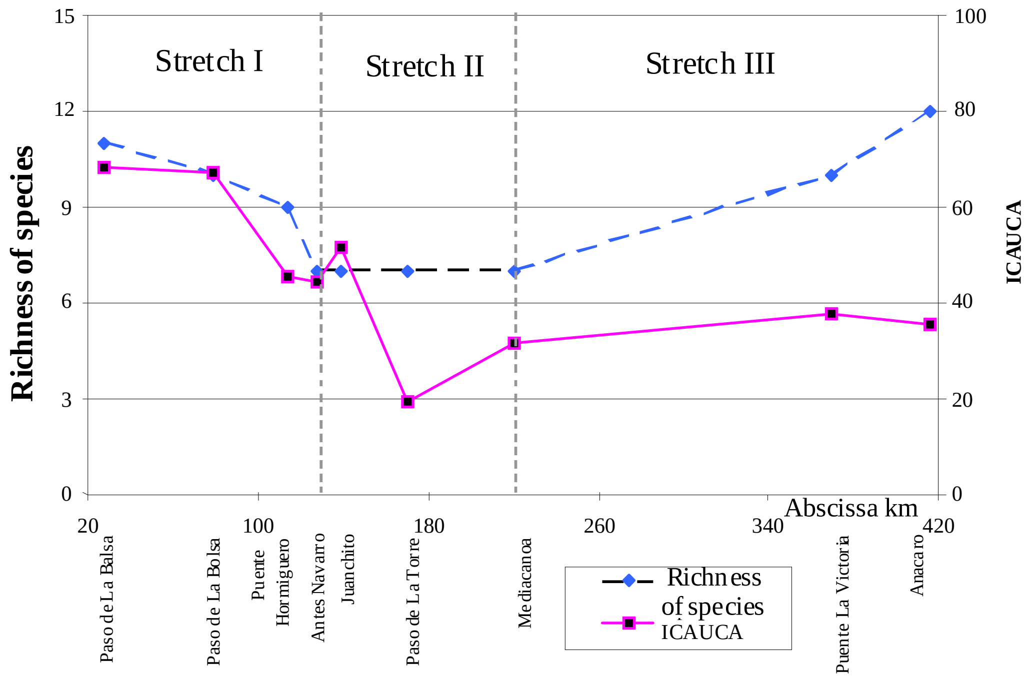


Figure 4.6 Variation of the richness of species (S) and the ICAUCA index throughout the Cauca river, during the monitoring campaign carried out during June of 2004. Source: CVC and Univalle (2004c).

4.1.3.3 Shannon Diversity (*H’*)

The diversity is the measure of the heterogeneity of a system, which refers to the amount and proportion of the different biologic elements which are present in the system. High

richness and diversity values are symptoms of healthy environments, low diversity values on the other hand, are sign of habitat deterioration.

The Shannon-Wiener index (also called Shannon diversity index (*H'*)) (Shannon and Weaver, 1963), is based on the assumptions that; (a) the sample collected is a random sample from an infinitely large population and (b) all families present in the community are represented in the sample. The objective aims at evaluating the community structure with respect to occurrence of species. This index relates the number of observed species (richness) to the number of individuals (abundance). The *H'* index, applied worldwide on a wide array of biological communities, is calculated with the formula:

$$H' \quad \sum_{i \ 1}^{S}\left[\frac{n_i}{n} * \ln\left(\frac{n_i}{n}\right)\right] \qquad (4.1)$$

where s is the number of species present in the sample; n_i is the number of individuals in the sample i of the community and; n is the total number of individuals in the sample. For classification purposes of the water quality, the *H'* index has been classified in 3 ranges: quite clean ($3<H'<5.0$); half polluted ($1.5<H'<3$); and quite polluted ($0.0 < H'< 1.5$) (Roldán G, 1992).

The Figure 4.7 shows a high positive correlation degree between the *H'* index and the ICAUCA index throughout the Cauca river, since both indices increased and decreased with the same form at the same stretches. The lowest *H'* and ICAUCA index were presented at the second stretch of the Cauca river, which is the zone that receives the largest polluted discharges, from the city of Cali and the industrial zones of Yumbo and Palmira cities, in contrast, the highest *H'* and ICAUCA index were presented at the first stretch of the river, were good water quality is present.

4.1.3.4 BMWP

The Biological Monitoring Working Party (BMWP) used in this research is an adaptation of the original BMWP (Armitage *et al.*, 1983) for rivers located at the Colombian southwest region, where the Cauca river is located, and was developed by Zúñiga (2009). For the BMWP index calculation, certain macroinvertebrate families are scored according to their sensitivity to organic pollution. Values assigned are between 10 (most sensitive taxa) to 1 (very tolerant ones). Annex D presents the values assigned to each

macroinvertebrate family according to their sensitivity to organic pollution used for the BMWP index calculation in the Cauca river. The BMWP is the total of the scores of all families present in a taxa list, where each family in the sample is counted only one time, regardless of the number of species. The final result of the BMWP index is a value, which can be expressed as well as a verbal description, a symbol or a colour. The water quality classification ranges according to the BMWP index can be seen in the Table 4.3.

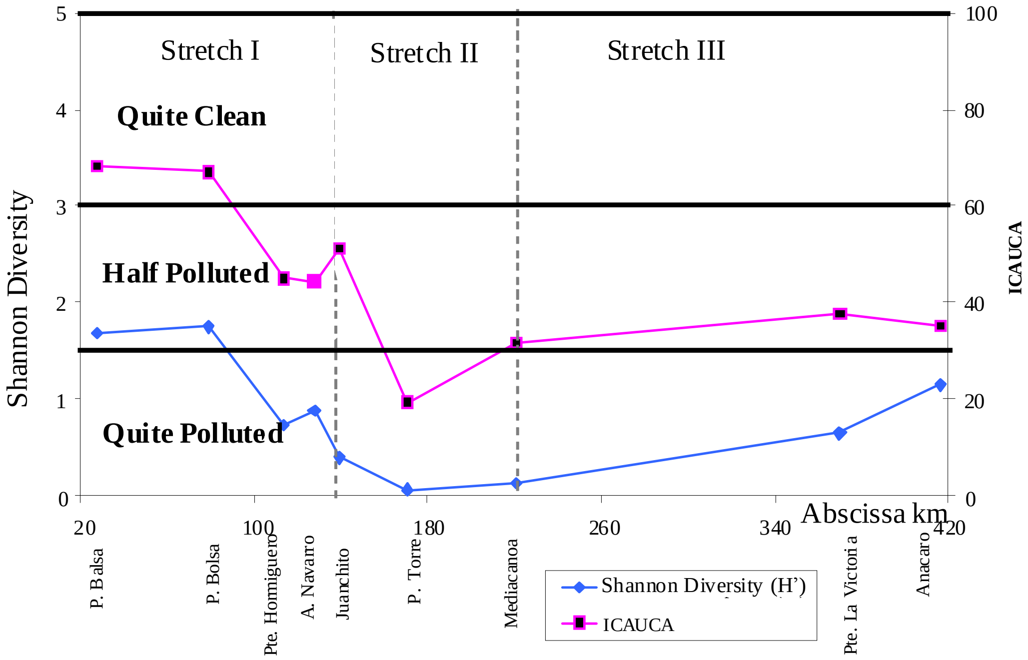


Figure 4.7 Variation of the Shannon Diversity (*H'*) and the ICAUCA index throughout the Cauca river, during the monitoring campaign carried out during June of 2004. Source: CVC and Univalle (2004c).

Table 4.3 Water quality classification ranges according to the BMWP adapted for rivers located at the Colombian southwest region, developed by Zúñiga (2009).

Class	Score	Category	Colour
	>120	Quite clean	Blue
I	100 - 120	No polluted	Blue
II	61 - 99	Start pollution	Green
III	36 - 60	Polluted	Yellow
IV	16 - 35	Quite Polluted	Orange
V	< 15	Septic	Red

Source : Zúñiga (2009)

The Figure 4.8 shows a high positive correlation degree between the BMWP adapted for rivers located at the Colombian southwest and the ICAUCA (Figure 4.3) index throughout the Cauca river, since both indices increased and decreased with the same form at the same stretches. In general, the BMWP values show that the water quality in the Cauca river is

already affected in the first stretch of the river, where the BMWP (i.e. 36 <ICAUCA<60) classifies the river as polluted. This impact could be related with the hydromorphological pressure generated by the Salvajina dam. Benthos communities below dams often show a reduction in species richness, while some species increase in abundance (Fruget, 1991). For example, invertebrate filter feeders (e.g., Hydropsychidae, Simulidae) often increase in numbers downstream of reservoirs (REBECCA, 2004). Afterwards, the BMWP shows progressive water quality deterioration, until it reaches a septic condition at the Juanchito station, in the second stretch, which is the most critical zone in terms of pollution. Finally, both the BMWP and the ICAUCA show a slight recuperation of the water quality at the third stretch, reaching a quite polluted water quality level in terms of the BMWP and an acceptable water quality level in the case of the ICAUCA.

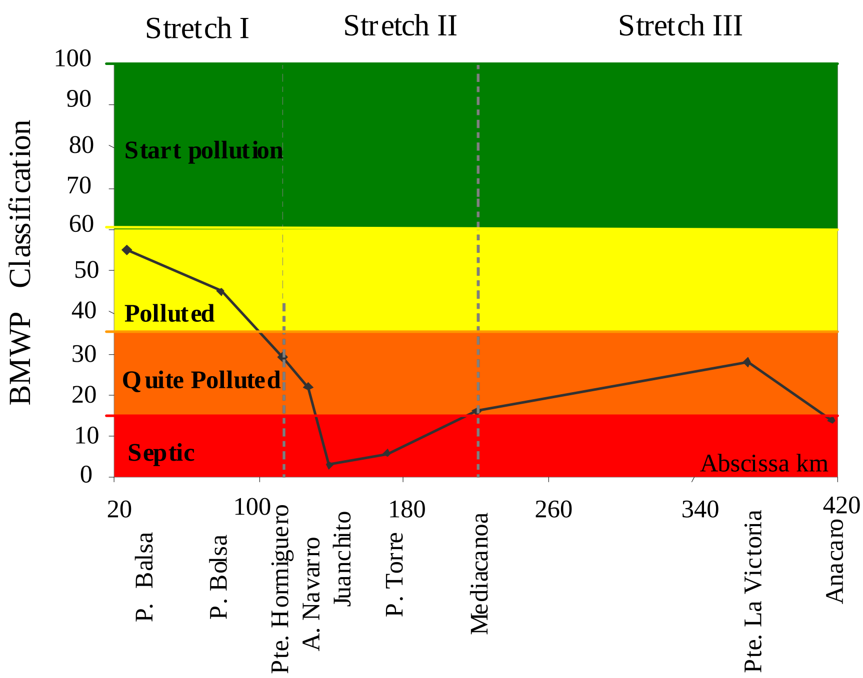


Figure 4.8 Variation of the BMWP and the ICAUCA index throughout the Cauca river, during the monitoring campaign carried out during June of 2004.Source: CVC and Univalle (2004c).

In spite of the similar behaviour observed between the biotic index BMWP and the physical-chemical index ICAUCA, the BMWP is more drastic in the water quality classification of the Cauca river. However, it is necessary to clarify that the BMWP used in this research has been developed for rivers in the Colombian southwest, rivers which have different conditions from those found in the Cauca river. For the Cauca river, it is important to consider, between other factors, the impact of the hourly fluctuation of the

domestic and industrial wastewater discharges into the river and the hourly river flow fluctuations generated by the operation of the Salvajina dam, which affects the dilution capacity of the river. Therefore, in order to have results with less uncertainty, it is necessary to perform in the future an adaptation of the BMWP for the specific conditions of the Cauca river.

4.2 WATER QUALITY MODELLING

The Cauca river is one of the most important rivers in Colombia. In its geographical valley, this river is used for numerous purposes: energy generation; water source for human consumption, agricultural and industrial use; mining of materials from the riverbed; fishing activities and recreation; but unfortunately this river is used as well as receiving aquatic ecosystems for industrial and domestic wastewater discharges from some municipalities located into the basin. In general, the environmental problem of water resource in the geographical valley of Cauca river is associated to degraded areas by inadequate use of land, raising in the water demand, some deforestation processes and erosion. One of the main problems is the degradation of the water quality, as consequence of the untreated wastewater discharges, the pollution generated by the mining exploitation and inadequate handling of solid waste. The regulations and other pollution control strategies have been supported by increase of the knowledge in the river behaviour. Taking into account the number of variables involved and the complex interrelations between these variables, mathematical modelling has been used as a planning and pollution control tool.

In the framework of the Cauca River Modelling Project (PMC), the software MIKE11 was used to simulate the hydrodynamic and water quality of the Cauca river. MIKE11 is a unidimensional and mechanistic dynamic model, in which inputs variables and state variables may vary with time, and thus result in a time variable output. Water quality modelling was carried out in Level 1 of MIKE11, which includes temperature, BOD, and DO as state variables. In order to perform the calibration and verification of the MIKE11 model, in the framework of the PMC project two water quality monitoring campaigns with continuous measurements, simultaneousness on different places and high sampling frequency were carried out during the months of August of 2003 (low flow conditions) and February of 2005 (high flow conditions). Taking into consideration that the reliability of the model depends on the calibration process, in this project was very important to have an

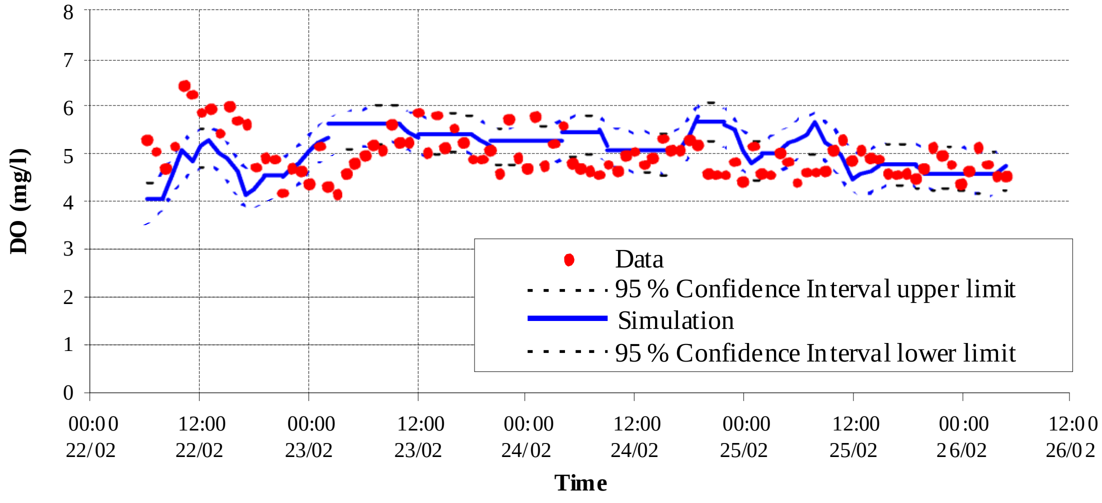
adequate calibration methodology that allowed identifying the set of parameters that fitted the output of the model with the observed data. Thus, a combination of skills of the modeller, together with a sensibility analysis and the technique called Generalized Likelihood Uncertainty Estimation-GLUE (Beven and Binley, 1992) were used during the calibration. The sensibility analysis allowed to identify the calibration parameters that most affected the answers of the model, whereas the GLUE technique, which is based on the method of controlled aleatory search of Monte Carlo, allowed to incorporate and to quantify the uncertainty that exists in the structure and the parameters of the model and the data observed with the purpose of determining the real predictive capacity of the model and to generate confidence intervals (CI) for the model results (95 % CI). During the calibration process of the water quality module of MIKE11 a total of 12 parameters were estimated.

The schematization of the Cauca river and tributaries used in the MIKE11 model can be seen in the Annex E. The Figure 4.9 shows the results of the calibration of the Cauca river water quality simulation model in one of the monitoring stations. Similar graphs were made for the rest of the stations and can be consulted at CVC, Univalle (2007b). In general, this model reproduces in acceptable form the values of DO, BOD_5 and temperature in the monitoring stations of the Cauca river, considering hourly fluctuations. This modelling approach allowed a very deep insight of the processes that occur in the river under dynamic conditions, such as temporary variations of flows and polluting loads.

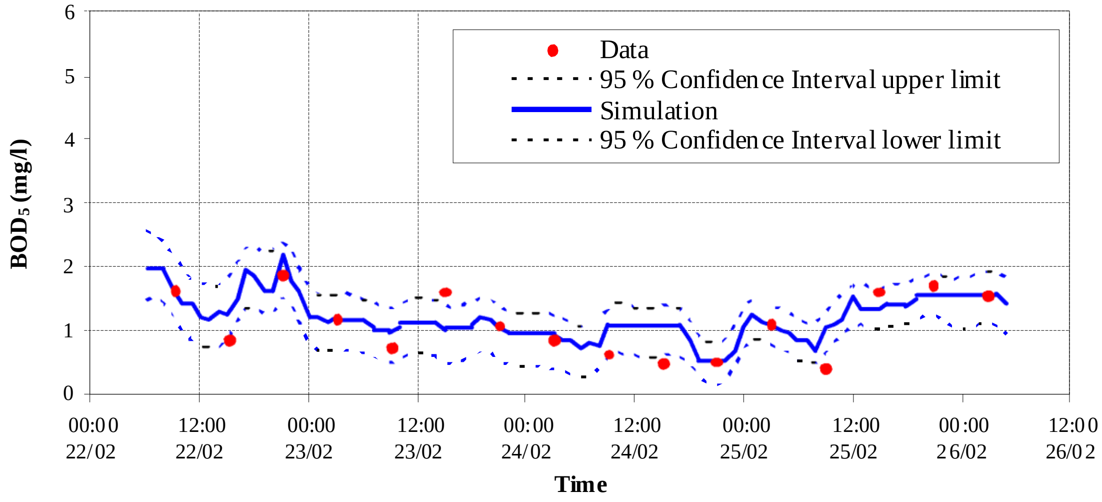
4.3 ECOLOGICAL MODELLING

The modelling approach followed in the framework of the PMC project using the software MIKE11, allowed a very deep insight of the water quality processes that occur in the river under dynamic conditions, however, this modelling approach just considered physical-chemical parameters, therefore biological components of the aquatic ecosystem were not taken into account. Thus, the development of an ecological modelling approach that integrates hydrodynamic, physical-chemical, and biological components is necessary in order to have a robust, reliable and effective tool to support river management and water policy in the Cauca river.

(a) DO Profile



(b) BOD_5 Profile



(c) Temperature profile

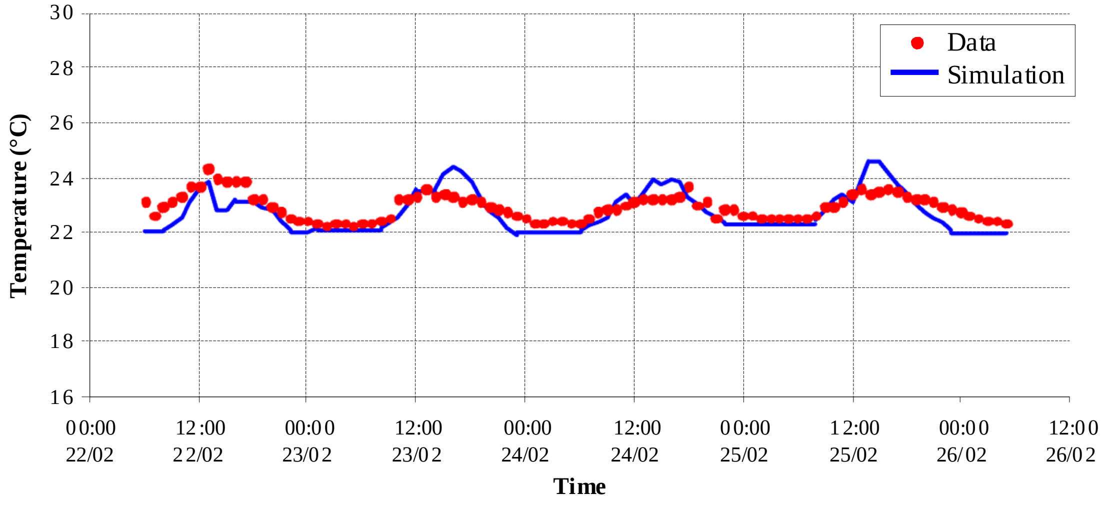


Figure 4.9 Calibration for the Cauca river water quality model at the station Juanchito. Simulation period: 22 of February – 26 of February of 2005.Condition: High flows. Source: CVC, Univalle (2007b).

The MIKE11 model for the Cauca river allows to calculate water quality variables such as temperature, BOD_5, and DO and hydrodynamic variables such as flow, depth and velocity. Taken into consideration the aim of this research, which is to build predict models for macroinvertebrate communities present in this river under different conditions by means of an integration with the MIKE11 model, the ecological statistical methods proposed in this research have to include these variables (temperature, BOD_5, DO, flow, depth and velocity) during its construction process.

In order to apply GLM regression analysis for predicting macroinvertebrates in the Cauca river, it was necessary first to construct a database with the monitoring information of the years 1997 (February), 2001 and 2004 (Annex. F Table F1). Aquatic macroinvertebrate identifications at species-level are costly and time-consuming and identification at higher taxonomic levels, such as class, may give rise to less reliable predict models. Taken into consideration the aim of this research and the financial resources available for biological assessment in the CVC, it was decided to build macroinvertebrate predict models at the taxonomic level of order. Therefore, a total of three (3) macroinvertebrate orders were selected for constructing the models. Ephemeroptera and Trichoptera orders (pollution sensitive benthos, which belongs to Phylum Arthropoda and Class Insect) as biological indicators for good water quality conditions and Haplotaxida order (pollution tolerant benthos, which belongs to Phylum Annelida and Class Oligochaeta) as biological indicator for polluted water with high organic matter content (Photo 4.2).

4.3.1 Logistic regression models for predicting the occurrence of macroinvertebrates

In order to apply logistic regression analysis for predicting the (probability of) occurrence of macroinvertebrates in the Cauca river, it was necessary to transform the abundance database to a record of presence (1) or absence (0) (Annex. F Table F2). The Table 4.4 shows the physical-chemical, hydrodynamic and biological database used for calculating the logistic regression models for the target organisms (i.e. Ephemeroptera, Trichoptera and Haplotaxida). Two (2) logistic regression approaches were performed in this research with the statistical software XLSTAT version 2009 (Addinsoft, 2009). The simple logistic regression model (SLRM) and the multiple logistic regression model (MLRG), which differ in the number of environmental variables considered (i.e. the first one only considers

one (1) variable in the model). The second approach gave the best results in this research; however results of both methods are presented in this discussion.

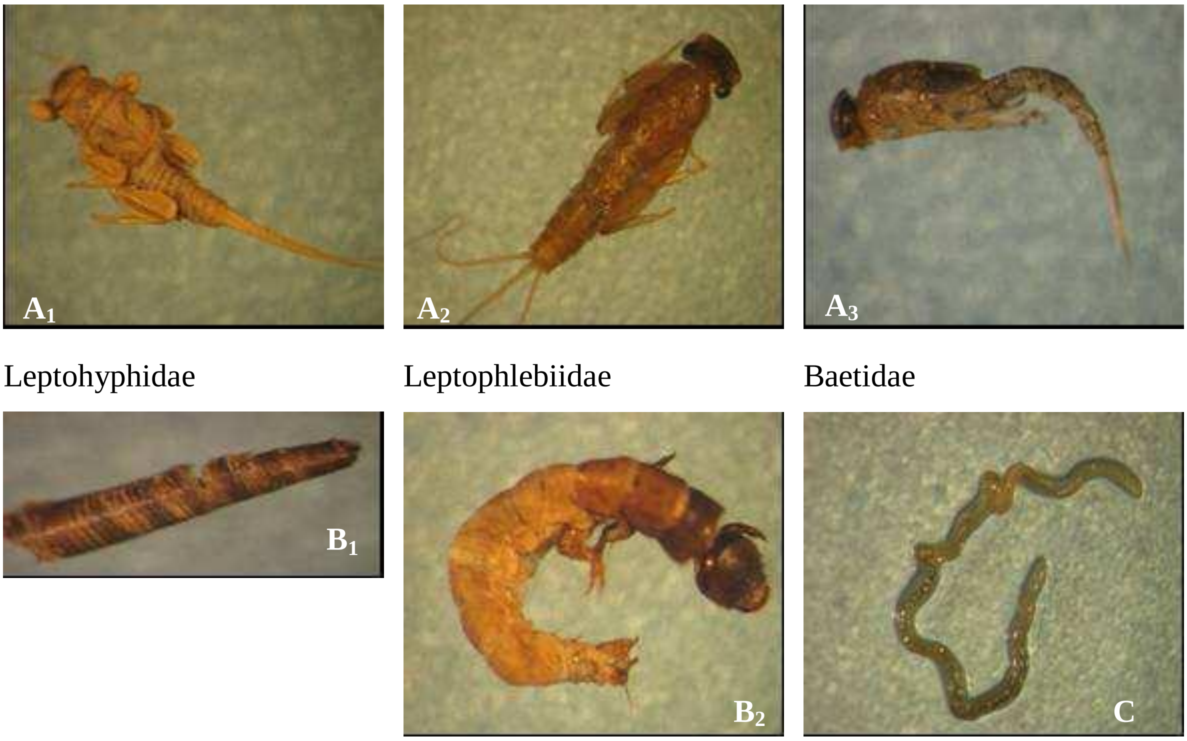


Photo 4.2 Macroinvertebrate orders selected for building predict models for biological communities at the Cauca river. A) Ephemeroptera (e.g. Leptohyphidae, Leptophlebiidae and Baetidae); B) Trichoptera (e.g. Leptoceridae and Hydropsychidae); C) Haplotaxida (e.g. Tubificidae).
Source: Photos taken from CVC and Univalle (2004c).

Table 4.4 Database used for calculating logistic regression models for predicting the probability of macroinvertebrate occurrence in the Cauca river.

Sampling site	Date	Condition	Item	Flow	Depth	Velocity	Temp	BOD_5	DO	Ephemeroptera	Trichoptera	Haplotaxida
Paso de la Balsa	19-Feb-97	High flow	1	269	4.9	1.2	18.0	3.7	6.8	1	1	0
Paso de la Bolsa	19-Feb-97	High flow	2	311	4.2	0.96	20.0	1.9	5.9	0	1	0
Juanchito	19-Feb-97	High flow	3	395	7.2	0.9	20.0	3.2	5.6	0	1	0
Paso de la Torre	19-Feb-97	High flow	4	419	6.9	0.9	20.0	4.6	4.5	0	0	1
Mediacanoa	19-Feb-97	High flow	5	509	6.3	0.9	22.0	6.8	3.0	0	0	1
Anacaro	25-Jul-01	Low flow	6	177	3.3	0.52	26.4	2.5	2.6	0	1	0
Paso de la Balsa	22-Jun-04	Low flow	7	83	3.4	0.62	24.0	0.2	6.9	1	1	0
Paso de la Bolsa	22-Jun-04	Low flow	8	109	2.8	0.55	22.0	0.1	6.7	1	1	0
Puente Hormiguero	22-Jun-04	Low flow	9	120	2.1	0.57	23.0	0.5	5.9	1	1	1
Antes Navarro	23-Jun-04	Low flow	10	128	5.0	0.58	25.0	1.4	5.1	1	0	1
Juanchito	23-Jun-04	Low flow	11	132	4.9	0.58	23.0	1.7	4.6	0	0	1
Paso de la Torre	23-Jun-04	Low flow	12	123	4.8	0.56	24.0	15.5	0.3	0	0	1
Mediacanoa	23-Jun-04	Low flow	13	152	3.7	0.55	23.0	12.8	0.8	0	0	1
Puente la Victoria	24-Jun-04	Low flow	14	165	2.4	0.55	24.0	3.3	2.3	1	0	1
Anacaro	24-Jun-04	Low flow	15	187	3.4	0.54	25.5	3.4	1.6	0	0	1

4.3.1.1 Simple logistic regression model

Simple logistic regression models (SLRM) for the target organisms (i.e. Ephemeroptera, Trichoptera and Haplotaxida) were performed considering each individual environmental variable (i.e. temperature, BOD_5, DO, flow, depth and velocity) separately. The Figure 4.10 presents an example of the results of SLRM for Trichoptera using as predictor variable DO. This figure shows a high positive correlation between DO concentrations and the probability of presence of Trichoptera, thus at high oxygen levels the probability of presence of Trichoptera is higher (e.g. at DO > 6 mg/l, the probability of presence is about 80 %). This analysis confirms the hypothesis that these organisms are pollution sensitive. In the Annex G. results of SLRM are presented considering the best correlations between environmental variables and target organisms.

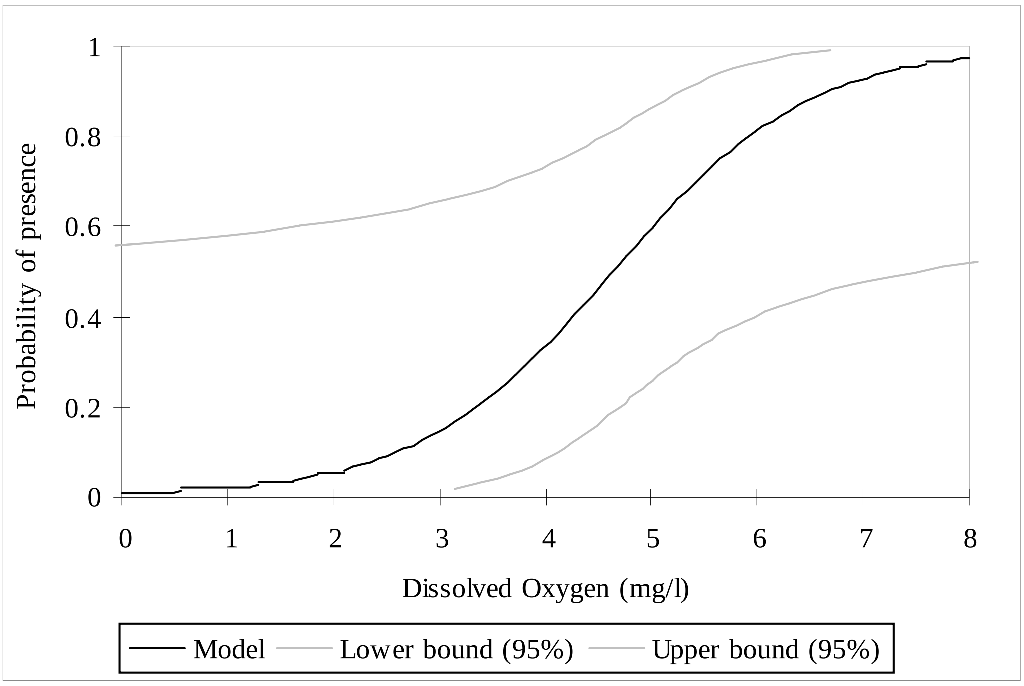


Figure 4.10 Results of SLRM for Trichoptera using as predictor variable dissolved oxygen

For each SLRM the software XLSTAT displays a large number tables and charts to help in analyzing and interpreting the results. In the following paragraphs the most important tables and charts for the results of the SLRM for Trichoptera using as predictor variable dissolved oxygen are presented. Similar analyses were performed for SLRM and MLRM for the rest of the target organisms (i.e. Ephemeroptera, Trichoptera and Haplotaxida):

1. Goodness of fit coefficients: this table displays a series of statistics for the independent model (corresponding to the case where the linear combination of explanatory variables reduces to a constant) and for the adjusted model (Table 4.5). The most important criteria are: the Akaike's information criterion (AIC), in which the best model is given by the lowest AIC value; -2 Log(Likelihood): the logarithm of the likelihood function associated with the model; the R^2(McFadden), R^2(Cox and Snell) and R^2(Nagelkerke) which are coefficients, like the R^2, between 0 and 1 which measures how well the model is adjusted (Addinsoft, 2009).

Table 4.5 Goodness of fit statistics : Trichoptera - SLRM with DO as predictor variable

Statistic	Independent	Full
Observations	15	15
Sum of weights	15.000	15.000
DF	14	13
-2 Log(Likelihood)	20.728	11.666
R^2(McFadden)	0.000	0.437
R^2(Cox and Snell)	0.000	0.453
R^2(Nagelkerke)	0.000	0.605
AIC	22.728	15.666
SBC	23.436	17.082
Iterations	0	6

2. Test of the null hypothesis H0: Y=p0: the H0 hypothesis corresponds to the independent model which gives probability p0 whatever the values of the explanatory variables (Addinsoft, 2009). The idea is to check if the adjusted model is significantly more powerful than this model. Three tests are available: the likelihood ratio test (-2 Log(likelihood)), the Score test and the Wald test (Table 4.6). The three statistics follow a Chi2 distribution whose degrees of freedom are shown. In this research the likelihood ratio test (-2 Log(likelihood)) was used as criteria and a set of significant threshold values (p-values) were fixed (i.e. $p < 0.001$ for a strong conclusion, $p < 0.01$ for a significant conclusion and $p < 0.05$ for a weakly but significant conclusion) to check the null hypothesis. For instance, if the likelihood ratio test is lower than 0.001, thus there is convincing evidence that the environmental variables included in the model are significant variables in predicting the target organism (Hosmer and Lemeshow, 2000). The results of this test shows that the SLRM for Trichoptera using as predictor variable DO gives a weakly but significant conclusion, therefore other approaches such as a different environmental variable or a MLRG could be evaluated.

Table 4.6 Test of the null hypothesis H0: Trichoptera-SLRM with DO as predictor variable

Statistic	DF	Chi-square	Pr > Chi²
-2 Log(Likelihood)	1	9.061	0.003
Score	1	7.253	0.007
Wald	1	4.179	0.041

3. Type III analysis: this table (Table 4.7) is only useful if there is more than one explanatory variable. Here, the adjusted model is tested against a test model where the variable in the row of the table in question has been removed. If the probability Pr > LR is less than a significance threshold which has been set (typically 0.05), then the contribution of the variable to the adjustment of the model is significant. Otherwise, it can be removed from the model (Addinsoft, 2009).

Table 4.7 Type III analysis : Trichoptera - SLRM with DO as predictor variable

Source	DF	Chi-square (Wald)	Pr > Wald	Chi-square (LR)	Pr > LR
Dissolved Oxygen	1	4.179	0.041	9.061	0.003

4. Equation of the model: the final regression model is then displayed by the software XLSTAT to make it easier to read or re-use the model. Thus the SLRM for predicting the probability of Trichoptera occurrence (*p*) in the Cauca river using as predictor variable dissolved oxygen is:

$$p_{(\text{SLRM, Trichoptera / DO})} \quad \frac{1}{1 \quad e^{(\;5 \quad 1.08*\text{DO})}} \qquad (4.2)$$

5. ROC curve: the Receiver Operating Characteristics - ROC curve (blue line in the Figure 4.11) is used to evaluate the performance of the model by means of the area under the curve (AUC) and to compare several models together. The terms used come from signal detection theory. The proportion of well-classified positive events is called the sensitivity. The specificity is the proportion of well-classified negative events. The curve of points (1-specificity, sensitivity) is the ROC curve. The AUC is a synthetic index calculated for ROC curves. The AUC, which ranges form zero to one, corresponds to the probability such that a positive event has a higher probability given to it by the model than a negative event (Addinsoft, 2009; Hosmer and Lemeshow, 2000). A model with good

discrimination ability is one that can correctly discriminate between occupied (presence) and unoccupied (absence) sites in an evaluation dataset. For a model with perfect discrimination, AUC=1 and for a random model (model with no discrimination ability), AUC = 0.5. A model is usually considered good when the AUC value is greater than 0.7. A well-discriminating model must have an AUC of between 0.87 and 0.9. A model with an AUC greater than 0.9 is excellent (Addinsoft, 2009; Pearce and Ferrier, 2000). Thus, the SLRM for predicting the probability of Trichoptera occurrence in the Cauca river using as predictor variable dissolved oxygen has excellent discrimination (AUC = 0.929 > 0.9).

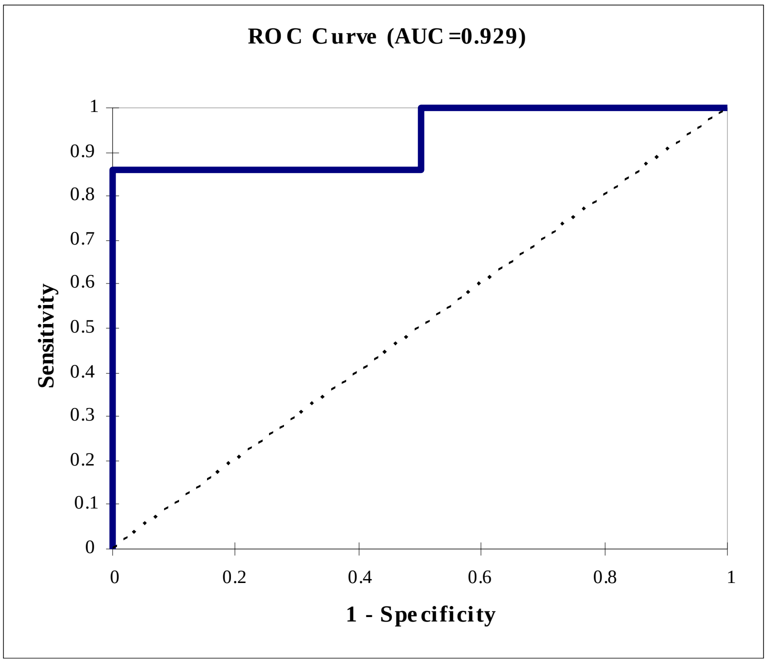


Figure 4.11 ROC curve for Trichoptera SLRM with DO as predictor variable

4.3.1.2 Multiple logistic regression models

A second logistic regression approach called multiple logistic regression model (MLRM) was performed in this research, which includes more than one independent environmental variable in the model. When developing models to predict community structure from environmental variables, an important problem is the selection of a limited number of environmental variables which will be used as predictors, because determining a large number of variables in the field is cost-inefficient, and furthermore, some variables may be redundant. Therefore, in order to select the most important environmental variables in the MLRM (combination of predictor variables) for predicting the probability of macroinvertebrate species occurrence in the Cauca river the Akaike's information criterion

(AIC) was used. The basis for comparing models was AIC values. The AIC can be used to determine Akaike weights for each model, which are the weights of evidence in favour of each model being the nearest to reality, given the other models being considered. MLRM were fitted using the maximum likelihood method (McCullagh & Nelder, 1989) with backwards elimination to select the final set of predictor variables. The step function, implemented in the statistical software XLSTAT version 2009 (Addinsoft, 2009) used in this research provides a procedure for this purpose using the AIC; this is a penalized version of the likelihood function in which the best model is given by the lowest AIC value. AIC is actually equivalent to twice the log-likelihood of the model fitted plus two times the number of parameters estimated in its formation (Rushton *et al.*, 2004). Given that the model with the smallest log-likelihood (lowest AIC value) is considered to be that with the best fit, the addition of two times the number of parameters means that AIC effectively includes a penalty for adding predictor variables to the model. The most obvious feature of this approach is that it is comparative, and leads to the identification of the best amongst a suite of models. Thus, the AIC criteria implemented in the software XLSTAT allowed identifying variables for inclusion or exclusion in models and to select the most parsimonious MLRM amongst a set (i.e. the MLRM with the lowest AIC value).

Tables 4.8, 4.9 and 4.10 present the results of the procedure implemented for selecting the best MLRM for predicting the probability of macroinvertebrate occurrence in the Cauca river. According to the AIC criterion, the best MLRM is that one with the lowest AIC value. The best MLRM for Ephemeroptera included DO, flow and depth as environmental predictor variable, the best MLRM for Trichoptera included the variables DBO_5, velocity, flow and depth and the best MLRM for Haplotaxida included DBO_5 and velocity as predictor variables. The final equations of the respective MLRM are presented in the equations 4.3, 4.4 and 4.5.

Table 4.8 Selection of the best MLRM for Ephemeroptera

MLRM for Ephemeroptera	**AIC**	**AUC**
DO, DBO_5, velocity, flow, depth, temperature	14.0	1.00
DO, DBO_5, velocity, flow, depth	12.0	1.00
DO, DBO_5, flow, depth	10.0	1.00
DO, flow, depth	**8.0**	**1.00**
DO, depth	14.8	0.93

$$p_{(\text{MLRM, Ephemeroptera})} \quad \frac{1}{1 \quad e^{(207.9 \quad 0.8\,Flow \quad 83.97\,Depth \quad 64.9\,DO)}} \tag{4.3}$$

Table 4.9 Selection of the best MLRM for Trichoptera

MLRM for Trichoptera	AIC	AUC
DO, DBO_5, velocity, flow, depth, temperature	14.0	1.00
DBO_5, velocity, flow, depth, temperature	12.0	1.00
DBO_5, velocity, flow, depth	**10.0**	**1.00**
DBO_5, flow, depth	15.0	0.95
DBO_5, flow	17.2	0.89

$$p_{(\text{MLRM, Trichoptera})} \quad \frac{1}{1 \quad e^{(121.64 \quad 1.04 Flow - 44.14 Depth \quad 103.4 velocity - 79.3 BOD_5)}} \tag{4.4}$$

Table 4.10 Selection of the best MLRM for Haplotaxida

MLRM for Haplotaxida	AIC	AUC
DO, DBO_5, velocity, flow, depth, temperature	22.09	0.963
DO, DBO_5, velocity, flow, temperature	22.09	0.963
DO, DBO_5, velocity, flow	19.05	0.944
DO, DBO_5, velocity	17.94	0.944
DBO_5, velocity	**17.61**	**0.926**

$$p_{(\text{MLRM, Haplotaxida})} \quad \frac{1}{1 \quad e^{(4.26 \quad 9.32\,velocity \quad 1.06\,BOD_5)}} \tag{4.5}$$

The complete outputs of the software XLSTAT for the best MLRM selected for predicting the probability of Ephemeroptera, Trichoptera and Haplotaxida occurrence in the Cauca river can be seen in the Annex H. The most important results obtained by applying the MLRM approach were:

- The results of the likelihood ratio test (-2 Log(likelihood)), for the null hypothesis for Ephemeroptera (-2 Log(Likelihood) = 0.0002) and Trichoptera (-2 Log(Likelihood) =

0.0004) had values lower than 0.001, therefore there is convincing evidence that the environmental variables included in the respective models are significant variables for predicting the target organisms. The likelihood ratio test for Haplotaxida had a value of 0.014 (which is lower than $p < 0.05$) therefore the results of this test gives a weakly but significant conclusion that the variables included in the model are significant for predicting Haplotaxida.

- The AUC for Ephemeroptera and Trichoptera had values of one (1), whereas the AUC value for Haplotaxida was 0.926, therefore the MLRM for predicting the probability of Ephemeroptera and Trichoptera can be classified as models with perfect discrimination, whereas MLRM for predicting Haplotaxida can be classified as excellent discriminating model. These means that the MLRM developed in this research can predict the occurrence (presence (1) or absence (0)) of Ephemeroptera, Trichoptera and Haplotaxida in the Cauca river appropriately.

4.3.2 Quasi-Poisson regression models for predicting the abundance of macroinvertebrates

Count data in ecology are often ‘‘overdispersed’’ (i.e. for any data set or model the variance exceeds the mean) therefore in order to use GLM techniques for estimating numbers of organisms (modelling macroinvertebrate abundance) at new sites and/or future times, in the Cauca river, the quasi-Poisson regression model (QPRM) approach was used instead of the Poisson regression approach in which the variance is equal to the mean. Table 4.11 shows the physical-chemical, hydrodynamic and biological database used for estimating the QPRMs for predicting the abundance of organisms for the target macroinvertebrates in the Cauca river (i.e. Ephemeroptera, Trichoptera and Haplotaxida). Annex I shows the statistical analysis made for the abundance data, where the evidence of overdispersion in the macroinvertebrate count data is presented (i.e. variance exceeds the mean). The regression coefficients (vector of parameters β) in the QPRM for modelling macroinvertebrate species abundance in the Cauca river were estimated by means of the maximum quasi-likelihood method using the statistical software S-PLUS version 6.1. The strategy followed for the explanatory variable selection in QPRM was to fit a sequence of models, beginning with a simple model with only an intercept term, and then include one additional explanatory variable in each successive model. The importance of the additional

explanatory variable was measured by the difference in deviances (or fitted log likelihoods) between successive models. Tables 4.12, 4.13 and 4.14 present the results of the procedure implemented for selecting the best QPRM for predicting the abundance of organisms for the target macroinvertebrates in the Cauca river. According to Residual Deviance (RD), the best QPRM is that one with the lowest RD. The best QPRM for Ephemeroptera included DO, velocity, flow and temperature as environmental predictor variables, the best QPRM for Trichoptera included the variables DO, depth, flow and temperature and the best QPRM for Haplotaxida included DO, depth and flow as predictor variables. The final equations of the respective QPRM are presented in the equations 4.6, 4.7 and 4.8. The complete outputs of the software S-PLUS for the best QPRM selected for predicting the abundance of Ephemeroptera, Trichoptera and Haplotaxida in the Cauca river can be seen in Annex J.

Table 4.11 Database used for estimating quasi-Poisson regression models for predicting the abundance of macroinvertebrates in the Cauca river.

Sampling site	Date	Condition	Item	Flow	Depth	Velocity	Temp	BOD_5	DO	Ephemeroptera	Trichoptera	Haplotaxida
Paso de la Balsa	19-Feb-97	High flow	1	269	4.9	1.2	18.0	3.7	6.8	27	10	0
Paso de la Bolsa	19-Feb-97	High flow	2	311	4.2	0.96	20.0	1.9	5.9	0	23	0
Juanchito	19-Feb-97	High flow	3	395	7.2	0.9	20.0	3.2	5.6	0	23	0
Paso de la Torre	19-Feb-97	High flow	4	419	6.9	0.9	20.0	4.6	4.5	0	0	53
Mediacanoa	19-Feb-97	High flow	5	509	6.3	0.9	22.0	6.8	3.0	0	0	58
Anacaro	25-Jul-01	Low flow	6	177	3.3	0.52	26.4	2.5	2.6	0	6	0
Paso de la Balsa	22-Jun-04	Low flow	7	83	3.4	0.62	24.0	0.2	6.9	9	7	0
Paso de la Bolsa	22-Jun-04	Low flow	8	109	2.8	0.55	22.0	0.1	6.7	12	3	0
Puente Hormiguero	22-Jun-04	Low flow	9	120	2.1	0.57	23.0	0.5	5.9	4	1	500
Antes Navarro	23-Jun-04	Low flow	10	128	5.0	0.58	25.0	1.4	5.1	4	0	752
Juanchito	23-Jun-04	Low flow	11	132	4.9	0.58	23.0	1.7	4.6	0	0	7510
Paso de la Torre	23-Jun-04	Low flow	12	123	4.8	0.56	24.0	15.5	0.3	0	0	32505
Mediacanoa	23-Jun-04	Low flow	13	152	3.7	0.55	23.0	12.8	0.8	0	0	6001
Puente la Victoria	24-Jun-04	Low flow	14	165	2.4	0.55	24.0	3.3	2.3	3	0	127
Anacaro	24-Jun-04	Low flow	15	187	3.4	0.54	25.5	3.4	1.6	0	0	195

Table 4.12 Selection of the best QPRM for Ephemeroptera

QPRM for Ephemeroptera	Residual Deviance
DO	248.98
DO, velocity	27.46
DO, velocity, temperature	25.42
DO, velocity, flow, temperature	**22.85**

$$_{\text{QPRM,Ephemeroptera}} \quad e^{12.636 \quad 0.022\,Flow \quad 3.096\,Velocity \quad 0.58\,Temp \quad 0.499\,DO} \qquad (4.6)$$

Table 4.13 Selection of the best QPRM for Trichoptera

QPRM for Trichoptera	Residual Deviance
DO	711.82
DO, depth	130.00
DO, depth, flow	95.00
DO, depth, flow, temperature	**39.30**

$$\text{QPRM,Trichoptera} \quad e^{58.69 \quad 0.0486\,Flow \quad 0.899\,Depth \quad 1.157\,Temp \quad 4.634\,DO} \tag{4.7}$$

Table 4.14 Selection of the best QPRM for Haplotaxida

QPRM for Haplotaxida	Residual Deviance
DO	23456
DO, depth	17784
DO, depth, flow	**12184**

$$\text{QPRM,Haplotaxida} \quad e^{7.912 \quad 0.017\,Flow \quad 0.985\,Depth \quad 0.479\,DO} \tag{4.8}$$

In order to validate the results of the predictive models for macroinvertebrate abundance data of the Cauca river, a comparison of the QPRM predictions with measured data considering the complete database (i.e. 15 observations) and subsets for high flow (i.e. the first 5 observations) and low flow (i.e. last 10 observations) conditions, were preformed following two approaches: (1) quantitative comparisons and (2) qualitative comparisons. Quantitative (or statistical) comparisons utilize statistical analyses to give quantitative measures of how good the model results fit the data. Statistical analyses provide a perspective on model – data comparison that numerically quantifies the state of the model calibration/verification. Statistical analyses such as the determination of the regression (R^2) are simple to apply and yield well - defined quantitative measures of model performance. The R^2 values vary from 0 to 1, thus the higher the R^2 the better the fitted model, however there are exceptional cases where the computational definition of R^2 can yield negative values, arise where the predictions which are being compared to the corresponding outcome have not derived from a model-fitting procedure using those data. Qualitative comparisons, on the other hand, are usually based on visual comparisons of the model

results with the data via time - series plots and spatial graphics for state variables, and then by determining whether the model reproduces observed patterns in time and in space.

In order to perform quantitative comparisons for the QPRM, the R^2 was used. Additionally, scatter plots of observed vs. predicted (OP graph) values (i.e. observed values in the ordinates vs. predicted values in the abscissas) was used to evaluate model predictions (Figures 4.12A; 4.13A and 4.14A). Piñeiro, *et al*. (2008) recommended this type of charts instead of the traditional predicted vs. observed (PO graph) to evaluate model performance. In order to perform qualitative comparisons, a profile of the observed data (abundance of each macroinvertebrate) and the model predictions were made, considering the complete database (Figures 4.12B; 4.13B and 4.14B) and subsets for high flow and low flow conditions (Figures 4.12C and D; 4.13 C and D and 4.14 C and D).

In general, the graphs show that the model reproduces with good precision the tendencies and the maximum and minimum values of abundance data for each macroinvertebrate (i.e. Ephemeroptera, Trichoptera and Haplotaxida) with high R^2 values ($0.866 < R^2 < 0.998$), with the exception of the datum of Anacaro station during the monitoring of low flow conditions of 2001 for Trichoptera (observation nr. 6 in the Figure 4.13 D, $R^2 = 0.42$) and the stations considered during the monitoring of high flow conditions of 1997 for Haplotaxida (Figure 4.13 C, R^2 with negative value). The behaviour presented in the datum of Trichoptera in the Anacaro station of 2001 could be associated to an outlier (i.e. an observation that is numerically distant from the rest of the data, that results from either sampling, measurement or estimation mistakes); whereas the behaviour of the model in the data of Haplotaxida during the monitoring of 1997 could be related with an instability of the model for predicting low abundance values of Haplotaxida. It is necessary to take into consideration that models are simplification of real systems, and per definition always contain errors in assumption, formulation and parameterization. In this case the presence of very high values of abundance during the monitoring of 2004 (i.e. values between 6001 and 32505 organisms) could affect the QPRM estimation, making that the model performs better at high abundance values rather than low abundance values. This analysis can be seen in the Figure 4.14 B, where the QPRM predictions fit very well at high abundance values of Haplotaxida, but at low abundance values the uncertainty in the predictions is higher.

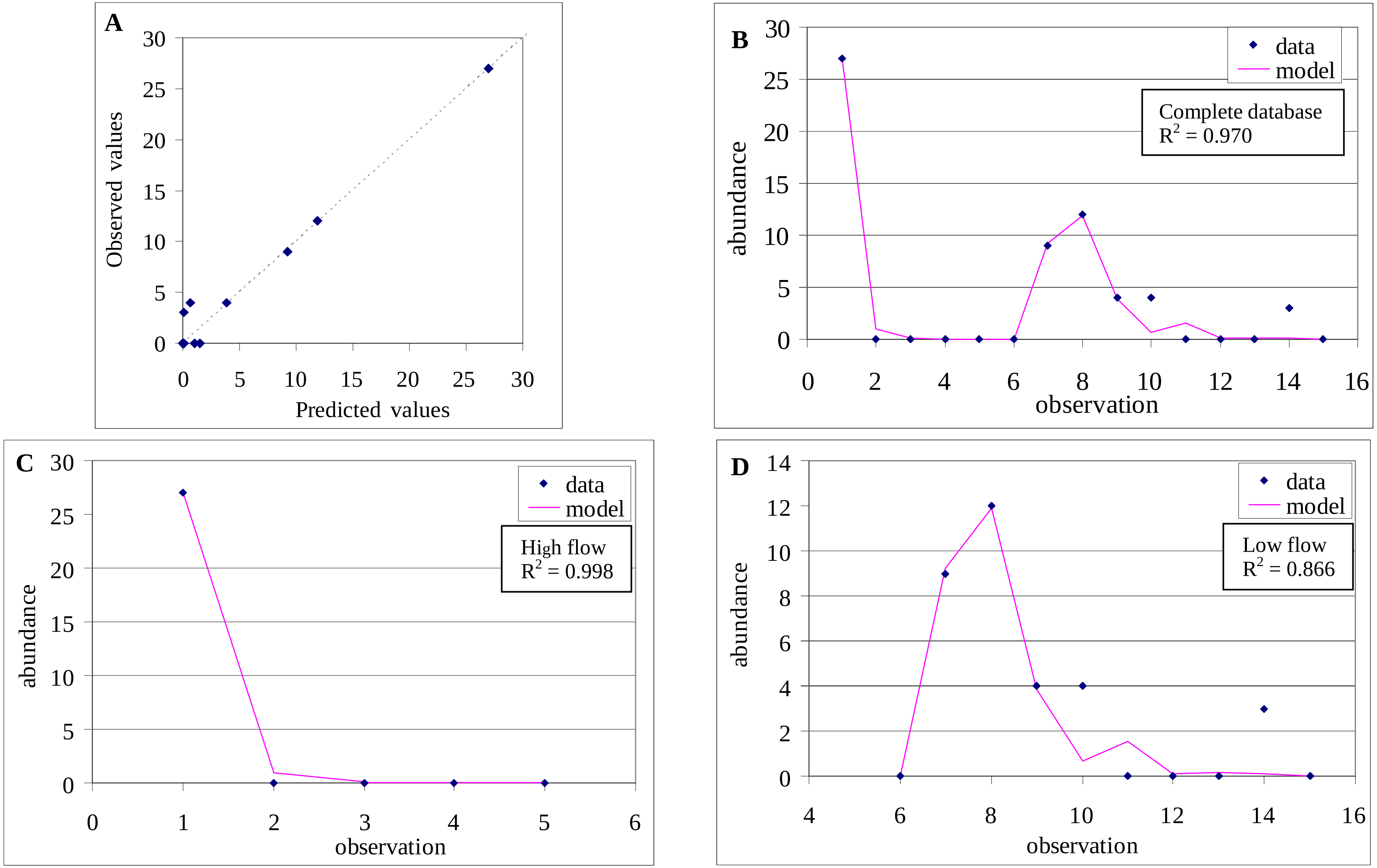


Figure 4.12 Results of the QPRM for predicting abundance of Ephemeroptera in the Cauca river

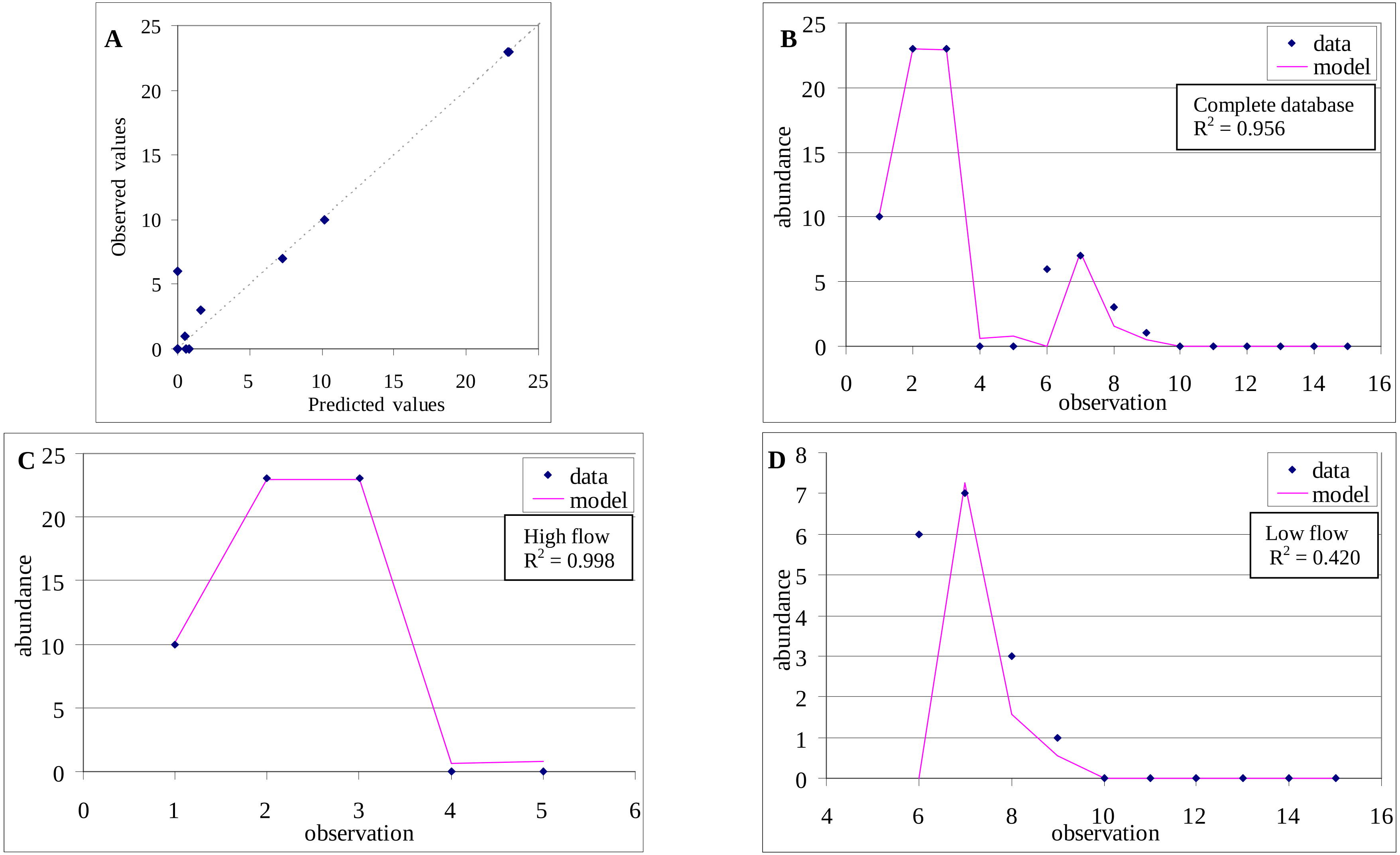


Figure 4.13 Results of the QPRM for predicting abundance of Trichoptera in the Cauca river

Chapter 4. Results

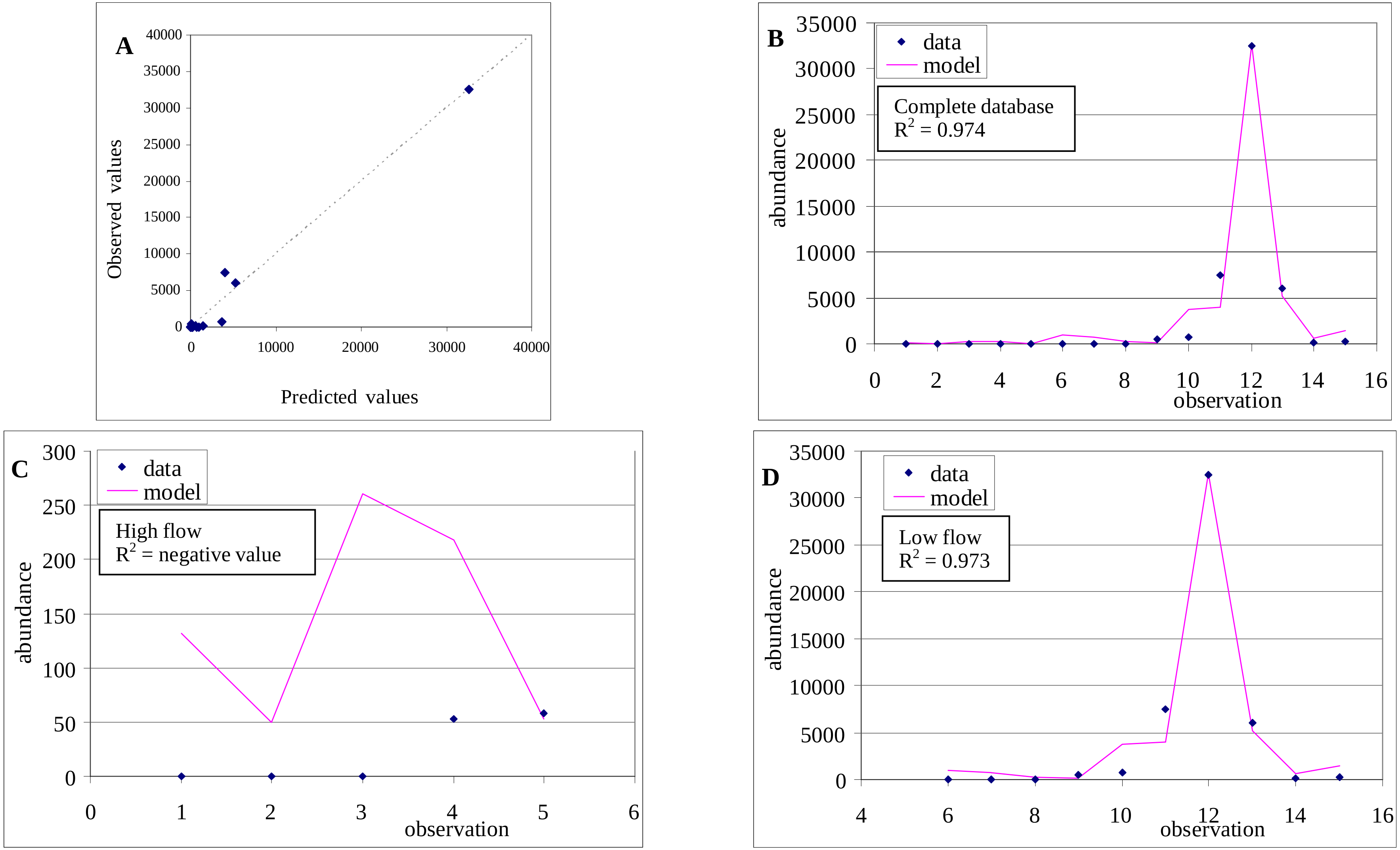


Figure 4.14 Results of the QPRM for predicting abundance of Haplotaxida in the Cauca river

Chapter 4. Results

4.3.3 Quasi-Poisson regression model for predicting the BMWP index

For developing a QPRM for the BMWP index, a similar procedure to that one done with the estimation of QPRM for macroinvertebrates was performed. Tables 4.15 presents the results of the procedure implemented for selecting the best QPRM for predicting the BMWP index in the Cauca river. The final equation QPRM for the BMWP index is presented in the equations 4.9. In general, the model reproduces with good precision the tendencies and the maximum and minimum values of the BMWP (Figure 4.15) with high a R^2 value (R^2 = 0.99). The complete outputs of the software S-PLUS for the best QPRM selected for predicting the BMWP index in the Cauca river can be seen in the Annex K.

Table 4.15 Selection of the best QPRM for the BMWP index

QPRM for Haplotaxida	**Residual Deviance**
DO	159.45
DO, Depth	125.34
DO, Depth, Velocity	82.92
DO, Depth, Velocity, Temperature	64.96
DO, Depth, Velocity, Temperature, Flow	**57.51**

$$_i \quad e^{3.639 \quad 0.002 Flow \quad 0.29 Depth \quad 2.089 Velocity \quad 0.229 Temp \quad 0.229 DO} \qquad (4.9)$$

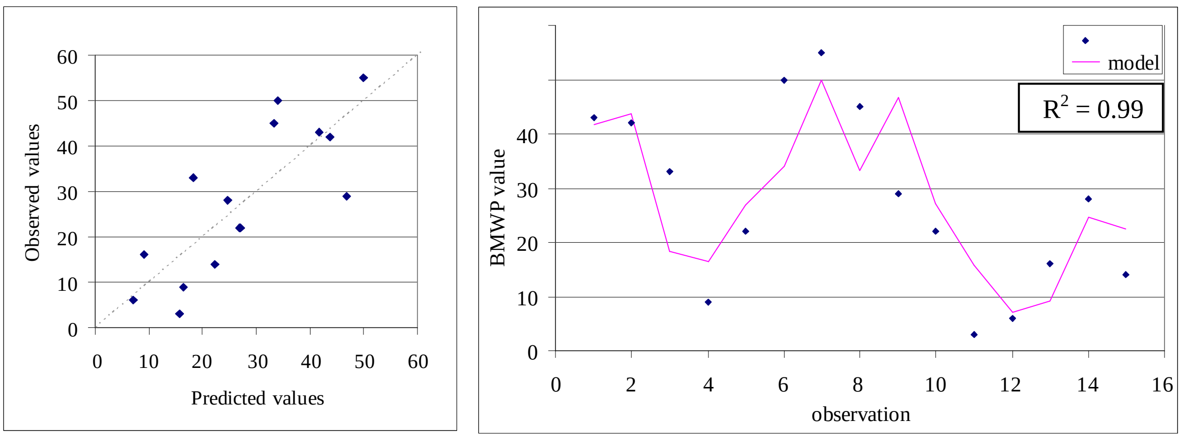


Figure 4.15 Results of the QPRM for predicting the BMWP index in the Cauca river

4.4 APPLICATION OF INTEGRATED ECOLOGICAL MODELLING OF THE CAUCA RIVER

With the model implemented in the framework of the PMC Project, and the models developed in this research for predicting the (likelihood of) occurrence and the abundance of macroinvertebrates in the Cauca river, some applications were carried out that allowed to study the effects in the water quality of the river generated by the plans and actions for the pollution control proposed by the environmental authorities, the municipalities and the industries in the Cauca river´s geographical valley.

The scenarios considered in this research correspond to the scenarios described by CVC and Univalle (2007a) for the year 2005 and 2015, considering the critical conditions for the dilution of the pollution in the river (i.e. low flow conditions). The scenarios considered were: reference situation scenario (E0-year 2005); very optimist scenario (E8-year 2015), very pessimistic scenario (E12-year 2015); and the scenario of water quality objectives proposed by the environmental authority CVC (E25-year 2015). The total organic matter load in terms of BOD_5 discharged into the Cauca river's basin considered in each scenario are: 204 Ton/day of BOD_5 in the reference situation scenario (E0), 106 Ton/day of BOD_5 in the very optimist scenario (E8), 255 Ton/day of BOD_5 in the very pessimistic scenario (E12); and 172 Ton/day of BOD_5 in the water quality objectives scenario (E25). Details of the database used to build the scenarios can be consulted in CVC and Univalle (2007a).

The Figure 4.16 presents the results of the four (4) scenarios considered in this research for the Cauca river obtained by means of the physicochemical and hydrodynamic model MIKE11. A profile of dissolved oxygen and BOD_5 predictions is presented for each scenario. In general these graphs shows that in spite of the reduction of the total organic matter load in terms of BOD_5 discharged into the Cauca river's basin considered in these scenarios, the DO concentrations in the station Paso de La Torre (in the most critical zone in terms of pollution, between the stations Hormiguero and Mediacanoa) never reach values higher than 2.6 mg/l, therefore these DO values are still lower than the minimum standard value established by the Colombian Decree 1594/84 for different uses of the water resource, which means, smaller than 70% of the DO saturation concentration (5.2 mg/l for the Cauca river).

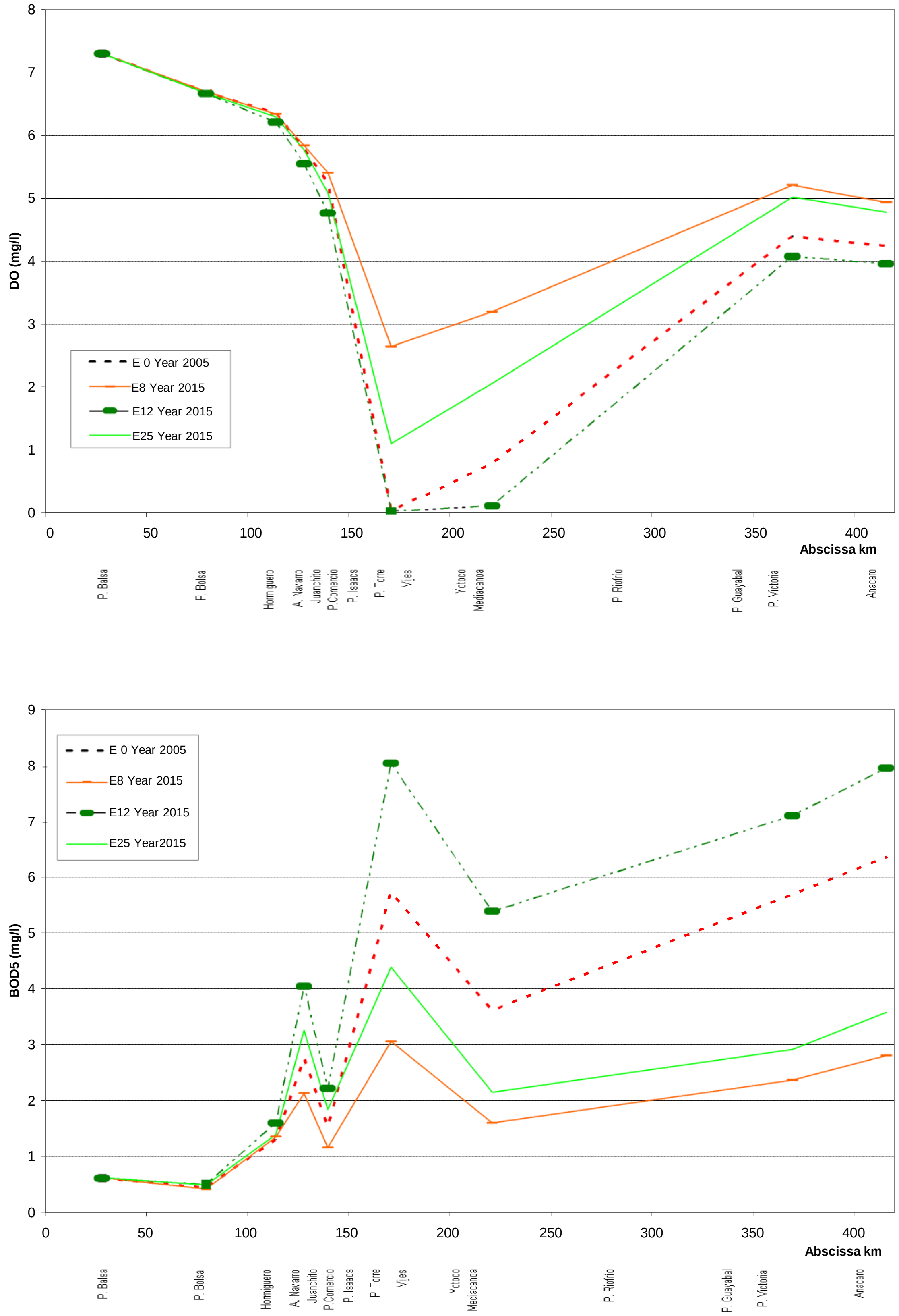


Figure 4.16 Dissolved Oxygen and BOD_5 predictions in the Cauca River for the pollution control scenarios considered in this research.

In order to apply the models developed in this research for predicting the probability of occurrence (logistic regression model) and the abundance (quasi-Poisson regression model) of macroinvertebrates, in the scenarios for the pollution control considered in the Cauca's river basin, a database with the physicochemical and hydraulic parameters calculated with the model MIKE11 for these scenarios was built (Table 4.16). The results of the application of the macroinvertebrate predictive models for the scenarios considered in the Cauca river can be seen in the Table 4.17.

The results presented in this table can be considered as a verification of the ecological models itself, because with this new database a sensitivity analysis of the models can be made. Thus, in the scenario with the highest pollution reduction (very optimist scenario, E8-year 2015) an improvement of the water quality of the Cauca river considering as reference situation the scenario E0 (year 2005) is achieved (increase in the number of cells with blue colour). This analysis can be illustrated with the increase of the number of pollution sensitive benthos (i.e. Ephemeroptera and Trichoptera) and a decrease of the number of pollution tolerant benthos (i.e. Haplotaxida). Additionally, in this scenario the values of the BMWP index increases in some stations indicating as well water quality improvement.

On the other hand, if the worst pollution condition scenario is considered (very pessimistic scenario, E12-year 2015) a deterioration of the water quality is obtained (increase in the number of cells with red colour). Thus, a decrease of the number of pollution sensitive benthos is observed and an increase of the number of pollution tolerant benthos is achieved. The BMWP index predictions for this scenario confirm the previous results showing a decrease in its values, indicating as well water quality deterioration.

Finally, if the scenario of water quality objectives proposed by the environmental authority is considered (E25-year 2015), which is an intermediate scenario in terms of pollution reduction, a water quality improvement is achieved in some stations, but there are other stations that show water quality deterioration (between Puente Hormiguero and Juanchito). This indicates that the pollution reduction proposed in the scenario is not enough for increasing the number of pollution sensitive benthos or decreasing the number of pollution tolerant benthos.

Table 4.16 Database used in the models for predicting macroinvertebrates, considered in the scenarios for pollution control in the Caucas' river basin

Reference situation scenario (E0-Year 2005)

Sampling site	Physical-chemical parameters			Hydraulic Parameters		
	DO (mg/l)	DBO_5 (mg/l)	Temp (°C)	Flow (m^3/s)	Depth (m)	Velocity (m/s)
Paso de La Balsa.	7.3	0.6	23.5	99.6	3.7	0.65
Paso de La Bolsa	6.7	0.4	21.5	125.9	3.0	0.61
Puente Hormiguero	6.4	1.3	22.5	120.6	2.1	0.57
Antes Navarro	5.8	2.8	24.5	129.7	5.0	0.58
Juanchito	5.3	1.5	22.5	133.9	4.9	0.58
Paso de La Torre	0.0	5.7	23.5	118.8	4.8	0.56
Mediacanoa	0.8	3.6	22.5	144.6	3.7	0.55
Puente La Victoria	4.4	5.7	23.5	155.0	2.3	0.55
Anacaro	4.2	6.4	25.0	176.5	3.4	0.54

Very optimist scenario (E8 -Year 2015)

Sampling site	Physical-chemical Parameters			Hydraulic Parameters		
	DO (mg/l)	DBO_5 (mg/l)	Temp (°C)	Flow (m^3/s)	Depth (m)	Velocity (m/s)
Paso de La Balsa.	7.3	0.6	23.5	99.6	3.7	0.65
Paso de La Bolsa	6.7	0.4	21.5	125.9	3.0	0.61
Puente Hormiguero	6.4	1.3	22.5	120.6	2.1	0.57
Antes Navarro	5.9	2.1	24.5	129.7	5.0	0.58
Juanchito	5.4	1.1	22.5	133.9	4.9	0.58
Paso de La Torre	2.6	3.1	23.5	118.8	4.8	0.56
Mediacanoa	3.2	1.6	22.5	144.6	3.7	0.55
Puente La Victoria	5.2	2.4	23.5	155.0	2.3	0.55
Anacaro	4.9	2.8	25.0	176.5	3.4	0.54

Very pessimistic scenario (E12-Year 2015)

Sampling site	Physical-chemical Parameters			Hydraulic Parameters		
	DO (mg/l)	DBO_5 (mg/l)	Temp (°C)	Flow (m^3/s)	Depth (m)	Velocity (m/s)
Paso de La Balsa.	7.3	0.6	23.5	99.6	3.7	0.65
Paso de La Bolsa	6.7	0.4	21.5	125.9	3.0	0.61
Puente Hormiguero	6.2	1.6	22.5	120.6	2.1	0.57
Antes Navarro	5.6	4.1	24.5	129.7	5.0	0.58
Juanchito	4.8	2.2	22.5	133.9	4.9	0.58
Paso de La Torre	0.0	8.1	23.5	118.8	4.8	0.56
Mediacanoa	0.1	5.4	22.5	144.6	3.7	0.55
Puente La Victoria	4.1	7.1	23.5	155.0	2.3	0.55
Anacaro	4.0	8.0	25.0	176.5	3.4	0.54

Scenario of water quality objectives (E25 - Year 2015)

Sampling site	Physical-chemical Parameters			Hydraulic Parameters		
	DO (mg/l)	DBO_5 (mg/l)	Temp (°C)	Flow (m^3/s)	Depth (m)	Velocity (m/s)
Paso de La Balsa.	7.3	0.6	23.5	99.6	3.7	0.65
Paso de La Bolsa	6.7	0.4	21.5	125.9	3.0	0.61
Puente Hormiguero	6.3	1.4	22.5	120.6	2.1	0.57
Antes Navarro	5.8	3.3	24.5	129.7	5.0	0.58
Juanchito	5.1	1.8	22.5	133.9	4.9	0.58
Paso de La Torre	1.1	4.4	23.5	118.8	4.8	0.56
Mediacanoa	2.0	2.1	22.5	144.6	3.7	0.55
Puente La Victoria	5.0	2.9	23.5	155.0	2.3	0.55
Anacaro	4.8	3.6	25.0	176.5	3.4	0.54

Table 4.17 Results of the application of the macroinvertebrate predictive models for the scenarios considered for pollution control in the Cauca river's basin.

Reference situation scenario (E0-Year 2005)

Sampling site	Logistic Regression Model			Quasi-Poisson Regression Model			
	Ephemeroptera	Trichoptera	Haplotaxida	Ephemeroptera	Trichoptera	Haplotaxida	BMWP
Paso de La Balsa.	1	1	0	12	52	546	51
Paso de La Bolsa	1	1	0	14	2	234	34
Puente Hormiguero	1	1	1	6	2	125	46
Antes Navarro	1	0	1	1	0	2461	29
Juanchito	1	0	1	3	0	2748	16
Paso de La Torre	0	0	1	0	0	39452	6
Mediacanoa	0	0	1	0	0	5850	8
Puente La Victoria	1	0	1	1	0	222	36
Anacaro	1	0	1	0	0	481	37

Very optimist scenario (E8 -Year 2015)

Sampling site	Logistic Regression Model			Quasi-Poisson Regression Model			
	Ephemeroptera	Trichoptera	Haplotaxida	Ephemeroptera	Trichoptera	Haplotaxida	BMWP
Paso de La Balsa.	1	1	0	12	52	546	51
Paso de La Bolsa	1	1	0	14	2	234	34
Puente Hormiguero	1	1	1	6	2	125	46
Antes Navarro	1	0	1	1	0	2391	29
Juanchito	1	1	1	3	0	2545	17
Paso de La Torre	0	0	1	1	0	11175	11
Mediacanoa	0	1	1	1	0	1844	14
Puente La Victoria	1	1	1	1	0	149	43
Anacaro	1	0	1	0	0	345	43

Very pessimistic scenario (E12-Year 2015)

Sampling site	Logistic Regression Model			Quasi-Poisson Regression Model			
	Ephemeroptera	Trichoptera	Haplotaxida	Ephemeroptera	Trichoptera	Haplotaxida	BMWP
Paso de La Balsa.	1	1	0	12	52	546	51
Paso de La Bolsa	1	1	0	14	2	234	34
Puente Hormiguero	1	1	1	6	1	134	45
Antes Navarro	1	0	1	1	0	2757	27
Juanchito	0	0	1	2	0	3465	15
Paso de La Torre	0	0	1	0	0	39461	7
Mediacanoa	0	0	1	0	0	8166	7
Puente La Victoria	1	0	1	0	0	259	33
Anacaro	1	0	1	0	0	552	34

Scenario of water quality objectives (E25 - Year 2015)

Sampling site	Logistic Regression Model			Quasi-Poisson Regression Model			
	Ephemeroptera	Trichoptera	Haplotaxida	Ephemeroptera	Trichoptera	Haplotaxida	BMWP
Paso de La Balsa.	1	1	0	12	52	546	51
Paso de La Bolsa	1	1	0	14	2	234	34
Puente Hormiguero	1	1	1	6	2	129	46
Antes Navarro	1	0	1	1	0	2482	29
Juanchito	1	0	1	3	0	2969	16
Paso de La Torre	0	0	1	0	0	23474	8
Mediacanoa	0	0	1	0	0	3191	11
Puente La Victoria	1	1	1	1	0	164	41
Anacaro	1	0	1	0	0	371	41

Water quality improvement considering the reference situation scenario (E0) [cyan]
Water quality deterioration considering the reference situation scenario (E0) [red]

5. GENERAL DISCUSSION AND FURTHER RESEARCH

Mathematical models are widely applied in science. 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 hydro-morphologic and physical-chemical processes (Figure 5.1). This situation is occurring in the Cauca's river context, where the environmental authority (CVC) has many hydrometric stations with hourly database, a few automated measurement stations for continuous water quality monitoring and historical information mainly based on discrete monitoring campaigns, with time intervals of months, however biological information for river quality assessment has been hardly recollected during few years. Therefore, the development of mathematical models for predicting biological communities (the aim of this research) together with the study of biological indicator species are complementary tools for river quality assessment and contribute to the integrated water quality management of this river.

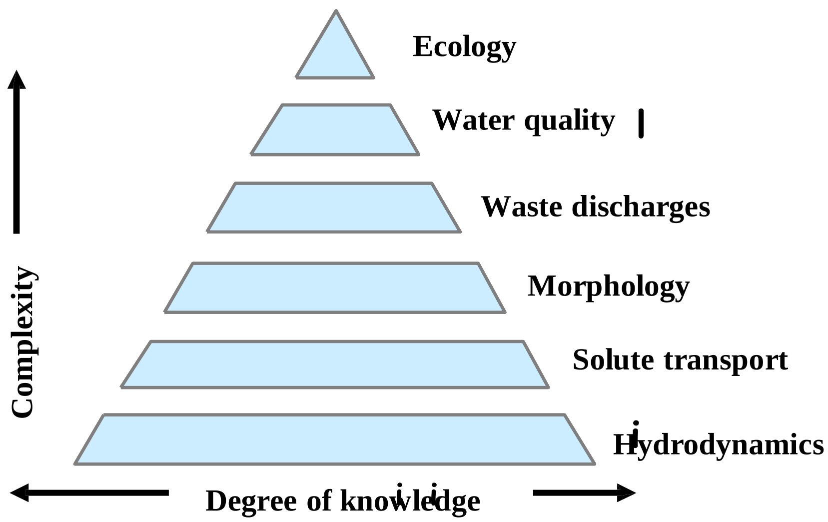


Figure 5.1 Knowledge of processes in ecosystems and its relation with ecological modelling. Source: Galvis (2004)

Assessment of the effect of human activities on river ecosystems requires indicators relating cause to effect (Figure 5.2). 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). This is the case of the Cauca river, which is affected by hourly fluctuations of the domestic and industrial wastewater discharges into the river and the operation of the Salvajina dam; the degradation of areas by inadequate use of land; the increase in the water demand and deforestation processes and erosion. These driving forces have been affecting the ecologic status of the river ecosystem, producing depletion of the dissolved oxygen and an increase of the abundance of pollutant tolerant benthos (e.g. Haplotaxida) and a decrease of pollution sensitive benthos (Ephemeroptera and Trichoptera), especially in the stretch located between the stations Puente Hormiguero and Mediacanoa, where the river receives high organic matter loads coming from domestic and industrial wastewater from the cities of Cali, Yumbo and Palmira.

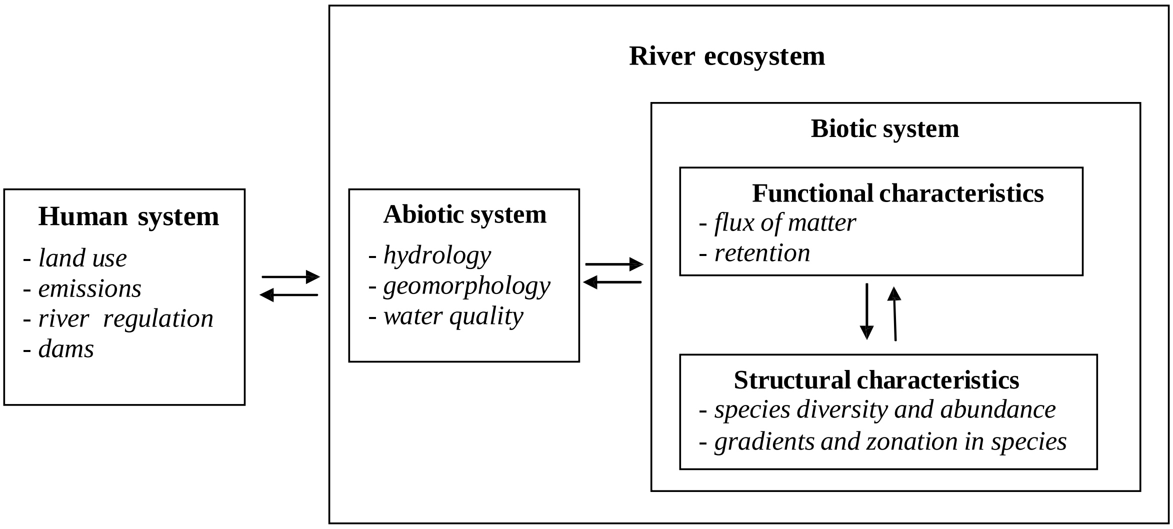


Figure 5.2 Cause–effect chain of the human system influencing the river ecosystem Source: Lorenz *et al*. (1997).

Nowadays, international legislation such as the European Water Framework Directive (WFD) focuses on the impacts of pressures on river biota which in turn could be used to assess river quality. The WFD requires that all European member states assess all their surface waters based on a number of biological elements, including macroinvertebrates. The aim of this Directive is to achieve a Good Ecological Status (GES) for all water bodies in the European Union by 2015. The ecological status refers to the quality of the structure

and functioning of aquatic ecosystem of the surface water. It is defined in terms of the quality of the biological community and the hydrological and chemical characteristics. In the United States, some regulatory agencies have begun to include macroinvertebrate habitat evaluations as a component of minimum flow evaluations. These evaluations have been made in recommendations for release schedules for hydropower facilities as part of re-licensing efforts with the Federal Energy Regulatory Commission and for allocating water abstraction rates for various municipalities. Most often, a suite of macroinvertebrate criteria has been used in the evaluations (Gore *et al*., 2001). Thus in the near future this approach could be used in the Colombian environmental legislation and therefore the development of mathematical models for predicting biological communities together with the study of biological indicators will be required to assess river quality.

5.1 INTEGRATED ECOLOGICAL MODELLING OF THE CAUCA RIVER

5.1.1 Application of integrated ecological modelling of the Cauca river

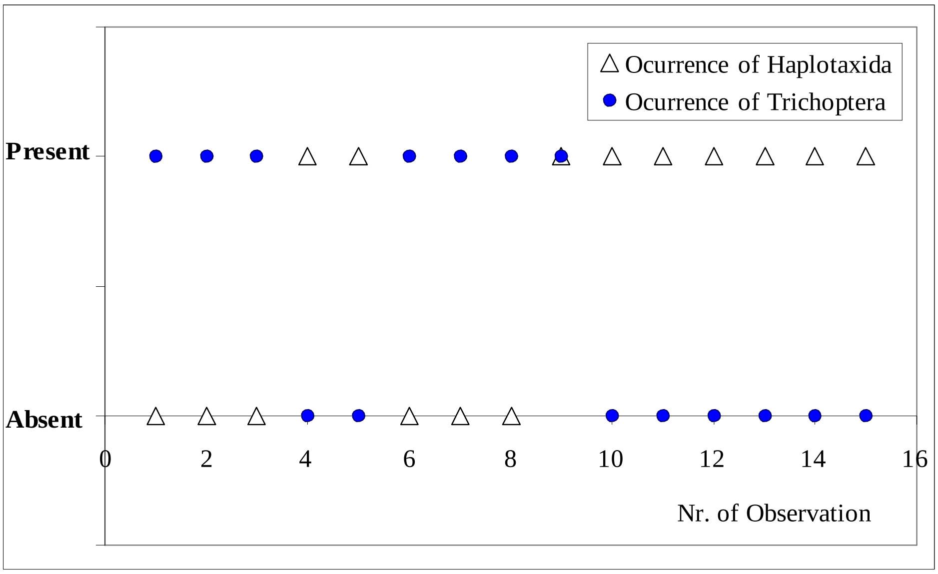
The enrichment of organic matter due to wastewater discharges in the zone near to the city of Cali affects flora and fauna living in the Cauca river mainly because the increased availability of food of a different quality changes the competitiveness of the different species in the river and because of a deoxygenation caused by increased heterotrophic activities. Further, the habitat of the riverbed can be changed into coverage of a bacterial film, and the oxygen status in the sediment can be deteriorated even without significant changes in DO in the river water (REBECCA, 2004). Severe deoxygenation may lead to a decrease of community diversity with only a few dominant species able to live under low oxygen conditions (e.g. Tubificidae). Mackenthun (1969) identified the following stepwise disappearance of macroinvertebrates subsequent to increasing pollution: stoneflies (Plecoptera), mayflies (Ephemeroptera), caddisflies (Trichoptera), scuds (Amphipoda), aquatic sowbugs (Isopoda), midges (Diptera) and bristle worms (Oligochaeta). Ammonium discharged or released by decomposition processes may be converted, at high pH and at high temperatures, to ammonia that in turn is extremely toxic to biota. Reduced oxygen concentration is one of the effects of organic matter enrichment since river biota can be impacted by organic matter even if the level of dissolved oxygen is unaffected (REBECCA, 2004).

Effects of organic matter degradation in streams are significantly influenced by local conditions of current velocity, substratum and channel morphology. When multiple impacts (e.g. habitat degradation or water pollution) are present, it is important to have aquatic habitat suitability specific assessment methods. An overwhelming scientific knowledge exists on the physical-chemical impacts of pollution of rivers with organic matter; however the major gap needed to fill is the lack of operational tools to be able to calculate the needed reductions in wastewater discharges of organic matter to meet biological quality criteria. This is the case of the Cauca river where the applications of the model MIKE11 performed in the framework of the PMC Project (CVC and Univalle, 2007a) allowed to study the effects in the water quality of the river in terms of DO and BOD_5 generated by the plans and actions for pollution control proposed by the environmental authority CVC, however, these scenarios never considered biological impacts on the aquatic ecosystem. Therefore, the proposal made in this research for the integration of ecological models for predicting the (likelihood of) occurrence and the abundance of macroinvertebrates for the Cauca river, together with the hydrodynamic and physical-chemical water quality model MIKE11, become a fundamental tool to calculate the needed reductions in wastewater discharges of organic matter to meet biological quality criteria in the Cauca river.

All modelling studies have three basic components: a data set describing the incidence or abundance of the species of interest and a data set of explanatory variables; a mathematical model that relates the species data to the explanatory variables; and an assessment of the utility of the model developed in terms of a validation exercise or an assessment of model robustness (Rushton *et al*., 2004).

Taken into consideration the aim of this research, which is to build predict models for macroinvertebrates communities present in this river under different conditions by means of an integration with the MIKE11 model, the ecological statistical models proposed in this research included during its calculation a data set of explanatory variables which could be calculated using the model MIKE11 (i.e. temperature, BOD_5, DO, flow, depth and velocity). An important consideration for all types of models is which and how many explanatory variables should be included in the model. If there are too few variables, the model will not be able to explain much of the variation. On the other hand, if there are too many variables, then the model will be too specific for the current data set (NIVA, 2007).

In this research, multiple logistic regression models (MLRM) for predicting the occurrence and quasi-Poisson regression models (QPRM) for predicting the abundance of Ephemeroptera, Trichoptera and Haplotaxida were built for the Cauca river. The Figure 5.3 shows that the macroinvertebrates selected for this ecological assessment are complementary biological indicators for polluted (Haplotaxida) and clean waters (Ephemeroptera and Trichoptera), because their geographic distribution in the Cauca river (presence or absence) depends on the their pollution tolerance. From this graph can be seen that Haplotaxida and Trichoptera were the most complementary organisms, thus when Haplotaxida is present, Trichoptera is absent and vice versa.



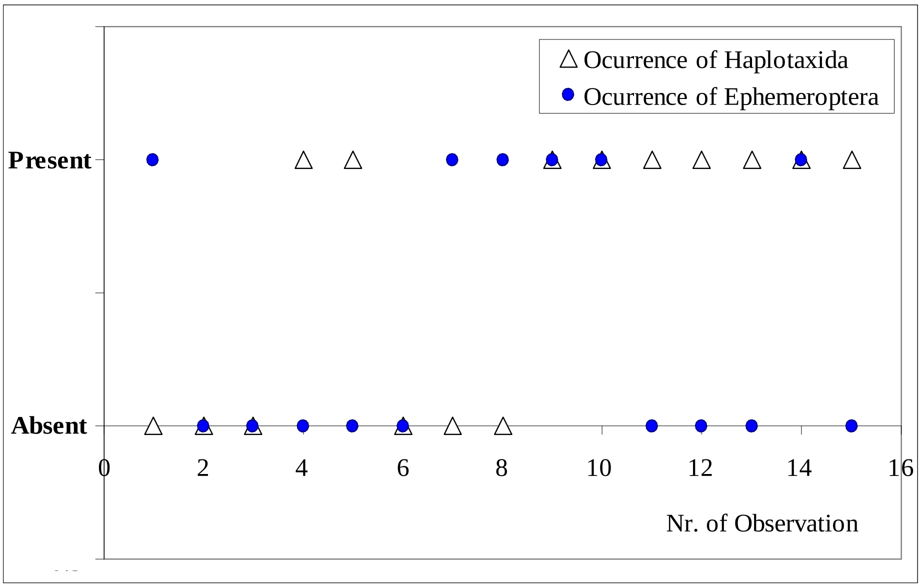


Figure 5.3 Macroinvertebrates selected for the development of predictive ecological models in the Cauca river as complementary biological indicators.

In order to study the importance of the different environmental variables estimated in the regression models developed in this research, all the values of the predictor variables in the MLRM and QPRM were compared (Table 5.1). From this table can be seen that flow velocity was the variable most important (cells with yellow colour) in four (4) of the seven (7) models developed (i.e. QPRM for Ephemeroptera and MLRM for Trichoptera and Haplotaxida and QPRM for BMWP index), followed by the depth in two (2) models (in the MLRL for Ephemeroptera and QPRM for Haplotaxida) and DO in one (1) model (in the QPRM for Trichoptera). Additionally, can be seen that in five (5) of the seven (7) predictive models the DO and the BOD_5 as indicators of organic pollution, were the first (DO for QPRM for Trichoptera) or the second most important variables (DO for MLRM Ephemeroptera and QPRM for Haplotaxida) and (BOD_5 for MLRM for Trichoptera and Haplotaxida). These results are in accordance with the literature, which reports that other factors than water quality are also important determinants of benthic communities (Figure 5.4). Of these the related factors of current velocity and nature of the substratum are overriding ones determining the nature of the community, especially in relation to invertebrates (Goethals, 2005).

Table 5.1 Values of the environmental predictor variables in the MLRM and QPRM

Variable	Value of the predictor environmental variable						
	Ephemeroptera		Trichoptera		Haplotaxida		BMWP index
	MLRM	QPRM	MLRM	QPRM	MLRM	QPRM	QPRM
Flow	-0.82	-0.022	1.04	0.049		-0.017	0.002
Depth	-83.97		-44.14	-0.899		0.985	-0.290
Velocity		3.096	103.42		-9.32		2.089
Temp		-0.580		1.157			0.229
BOD_5			-79.30		1.06		
DO	64.91	0.499		4.634		-0.479	0.229

First most important (yellow)
Second most important (cyan)
Third most important (green)
Fourth most important (orange)
Fifth most important (red)

From the Table 5.1 can be seen as well that flow velocity has a positive impact (positive value) on the prediction of sensitive pollution organisms (i.e. the higher the velocity, the higher the abundance or the higher the probability of presence) whereas the water depth has negative impact on same organisms (i.e. the higher the water depth, the lower the abundance or the lower the probability of presence). These results can be explained with

the mathematical relationship between the velocity and water depth with the dissolved oxygen. For the Cauca river the empirical predictive model selected for representing the reaeration process (i.e. the process of the exchange of oxygen between the atmosphere and the water mass) in the MIKE11 model, was the equation of O'Connor y Dobbins:

$$K_a \quad 3.93\frac{U^{0.5}}{H^{1.5}} \qquad (5.1)$$

where, K_a= reaeration rate [d^{-1}], U= flow velocity [m/s] and H= water depth [m].

This equation explains that when the water depth increases, less oxygen transfer from atmosphere to the water column can be achieved. On the other hand, if the flow velocity increases, more oxygen is transferred from the atmosphere and therefore the DO in the water column increases. If conditions of low DO transfer processes (i.e. high water depth and low flow velocities) predominate in the water column, less oxygen will be available for species and therefore a decrease of community diversity with only a few dominant species able to live under low oxygen conditions will remain (e.g. Haplotaxida).

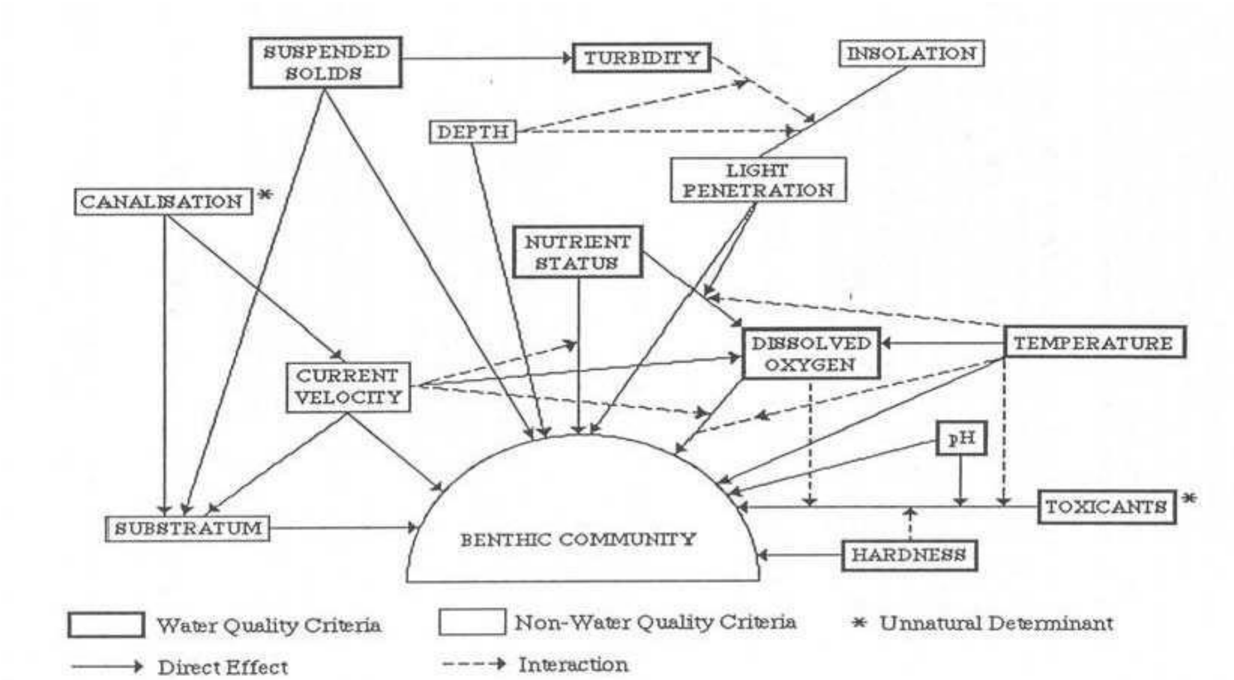


Figure 5.4 Water quality and non-water quality determinants of benthic communities in rivers. Source: De Pauw and Hawkes (1993), cited by Goethals (2005).

Regarding the application of the integrated ecological modelling of the Cauca river it was found that the MLRM and QPRM predict well the ecological impact of the scenarios for pollution control in the Cauca's river basin. Thus, in the scenario with the highest pollution reduction (very optimist scenario, E8-year 2015) an improvement of the water quality of the Cauca river is achieved, which is represented with the presence and/or an increase of the number of pollution sensitive benthos (i.e. Ephemeroptera and Trichoptera) and the absence and/or a decrease of the number of pollution tolerant benthos (i.e. Haplotaxida). On the other hand, if the worst pollution condition scenario is considered (very pessimistic scenario, E12-year 2015) a deterioration of the water quality is obtained, which is represented with the absence and/or a decrease of the number of pollution sensitive benthos and the presence and/or an increase of the number of pollution tolerant benthos. Additionally, the BMWP index predictions for these scenarios confirm the previous results, showing an increase of the BMWP values in some monitoring stations in the case of the very optimist scenario (E8) indicating water quality improvement and a reduction of the BMWP values in the case of the very pessimistic scenario (E12) indicating water quality deterioration.

5.1.2 Model reliability

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 (REBECCA, 2004). However, as few alternatives are available, they remain key tools for environmental quality assessment. The third component of ecological modelling is model assessment. This is not only an important check on the adequacy of the models developed, but it is an important factor determining their utility. 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 data set (Rushton *et al*., 2004). Evaluating the predictive performance of models is a vital step in model development. Such evaluation assists in determining the suitability of a model for specific applications. It also provides a basis for comparing different modelling techniques and competing models, and for identifying aspects of a model most in need of improvement (Pearce and Ferrier, 2000).

The simplest way to assess the reliability of any model is to compare model predictions with observed data. Assessment of multiple logistic regression models (MLRM) is complicated by the fact that the predictions from such models are proportions in the range 0–1 while most test data against which they may be compared comprise records of 1 or 0. In effect, model assessment involves comparing probabilities with categories. The approach followed in this research for the assessment of logistic regression models was the use of threshold-independent approaches such as Receiver Operating Characteristics (ROC plots). These are based on plotting the true positives against the false positive fractions for a range of thresholds in prediction probability. The area under the curve (AUC) for a ROC plot is taken as a measure of the accuracy of the model that is not dependent on a single threshold (Rushton *et al*., 2004). This AUC index is related with the discrimination capacity of the model (the ability of a model to correctly distinguish between occupied and unoccupied sites).

In order to assess the reliability of QPRM a predictive validation procedure, by means of comparison of the predictions with measured data considering the complete database and subsets for high flow and low flow conditions, was performed. Additionally, a sensitivity analysis of the both models (MLRM and QPRM) was performed considering a new database, corresponding to the scenarios for pollution control in the Cauca's river basin. Thus, the models were used predictively and not simply as a means of exploring relationships in the data set

The assessment of MLRM reliability showed that the models for Ephemeroptera (AUC=1), Trichoptera (AUC=1), and Haplotaxida (AUC=0.926) correctly discriminates between occupied (presence) and unoccupied (absence) sites in the dataset. Regarding the predictive validation procedure for QPRMs, it was found that in general the models reproduce with good precision the tendencies and the maximum and minimum values of abundance data for each macroinvertebrate (i.e. Ephemeroptera, Trichoptera and Haplotaxida) and the BMWP index, with high R^2 values ($0.866 < R^2 < 0.998$). However, it was found that the QPRM for Haplotaxida model performs better at high abundance values rather than low abundance values. The results presented in the application of the integrated ecological modelling of the Cauca river considering the scenarios for pollution control in the basin, can be understood as a verification of the ecological models itself, because with a new database (scenarios database) a sensitivity analysis of the models was performed. Thus, a

scenario which indicates a water quality improvement (very optimist scenario, E8), leads to an increase of the community diversity, with a raise of the abundance of pollution sensitive benthos (i.e. Ephemeroptera and Trichoptera) and a decrease of the number of pollution tolerant benthos (i.e. Haplotaxida). On the other hand a scenario which indicates a water quality deterioration leads to a decrease of the community diversity, with a raise of the abundance of pollution tolerant benthos and a reduction and/or disappearance of pollution sensitive benthos.

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 in the ecological models implies the identification of errors, inexactness, imperfection and unreliability in the models (Lek, 2007). The uncertainty of an ecological model is caused by both the lack of knowledge (i.e. data imperfection) and the variability of models and parameters (models' sensitivity) (Lek, 2007). Data may contain errors that result from either sampling, measurement or estimation mistakes. Analysed data are almost always incomplete with large and unknown amounts of measurement error or data uncertainty. Often the expense of data collection prohibits collecting as much data as might be desirable (Lek, 2007).

GLMs, developed in this research (i.e. MLRM and QPRM) 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). Therefore, the application of these models is limited to the specific geographical area where they were developed (i.e. the Cauca's river basin). Additionally, it is necessary to consider that ecological models are most reliable when applied within the range of observations used to construct the model. Extrapolation from empirical data is known to be uncertain. Therefore, this limitation has to be considered when applying these GLMs in the Cauca's river.

5.1.3 Ecological relevance of the model

Environmental decision-making is extremely complex due to the intricacy of the systems considered and the competing interests of multiple stakeholders. In the Cauca river´s geographical valley the public, industry, government and environmental authorities acknowledge that protecting, improving, and managing the water quality of the river is extremely important. During the last decade (1997-2007) in the framework of the Cauca River Modelling Project (PMC), researchers have responded by designing and implementing experiments, collecting and analyzing data, and developing and parameterizing models (MIKE11) in order to better understand and provide predictions about the physical-chemical water quality of the river. However, in the framework of this modelling approach has had a lack of mathematical tools to be able to calculate the needed reductions in wastewater discharges of organic matter to meet biological quality criteria. Therefore, the integration of ecological models for predicting the occurrence and the abundance of macroinvertebrates for the Cauca river, together with the model MIKE11, fills one of the major gaps that has the integrated water quality management of the Cauca river in Colombia.

5.2 COMPARISON WITH OTHER MODELS

The application of ecological modelling approaches that integrate hydrodynamic, physical-chemical, and biological components sub-models for predicting macroinvertebrates in rivers, is rather limited and hardly described in literature. One of the models available for this purpose is the Water Framework Directive Explorer (WFD-Explorer). The WFD-Explorer is being developed by a consortium of research institutes (Deltares) and water management authorities in the Netherlands. The WFD Explorer performs a number of functions related to the implementation process of the WFD in the European Union. The WFD Explorer allows the assessment of potential measures to achieve the environmental objectives. First of all it provides a platform to integrate and disclose ecological information on water bodies, including a common knowledge base on the effectiveness of measures. It also provides stakeholders with the possibility to deepen their understanding of the relationships between objectives pursued, measures that might be taken and impacts of such measures (Deltares, 2009).

The WFD-Explorer is a modular toolbox which supports integrated water management analysis in a river basin and it is mainly an appropriate tool for modelling the impact of different restoration measures on river ecology based on expert rules embedded in this simulation environment. The rules link the physical and chemical water body characteristics to an Ecological Quality Ratio (EQR) for the different aquatic communities (fish, macrophytes and macroinvertebrates) of each water body. The WFD-Explorer has a user interface for end users only, not for modellers. This tool has 47 pre-defined measures that must be specified by users, which afterwards are translated into steering variables. Finally, these steering variables by means of knowledge rules are transformed to aquatic ecological quality (i.e. EQR).

The WFD-Explorer has a database with ecological models and expert rules. The ecology rules are based on a large data set for all waterbodies in the Netherlands. In this mathematical tool a neural network is used to derive relations between variables considered in the model, such as: nutrient concentrations (N and P); oxygen and chloride concentrations: morphological parameters (shore protection/development, meandering/sinuosity); maintenance and shipping indicators; connectivity, water level dynamics, shadow class, weir/fish passage between others. The relations are used in a regression-tree model to derive critical pathways.

However, the WFD-Explorer has some limitations: 1. WFD-Explorer is a steady state application, thus not suitable for dynamics; 2. It simplifies water quality processes as a retention factor; 3. The knowledge rules for ecology are fixed for most users; 4. The main focus area is The Netherlands, hardly any applications outside The Netherlands.

Mouton *et al*. (2009b) evaluated the strengths and weaknesses of the Water Framework Directive Explorer (WFD-Explorer) toolbox on the Zwalm River basin in Flanders, Belgium. They found that in order to generate reliable information for supporting decision-making implementation of the WFD in the Zwalm River basin, a combination of the WFD-Explorer results with those of more detailed studies on physical habitat restoration impacts is needed. Thus, the spatial scale on which the WFD-Explorer modelled the impact of physical habitat restoration may have been too coarse to generate reliable results concerning such restoration measures.

5.3 FURTHER RESEARCH

There are other researches that could be developed in the future and can contribute to the integrated water quality management of the Cauca river in Colombia. These topics are discussed and suggestions how such research can be set up are presented.

5.3.1 Combination of mechanistic and datadriven models for predicting macroinvertebrates

In general, two approaches can be followed when performing (ecological) modelling which are mechanistic (i.e. food-webs) and data driven modelling (e.g. statistical models or artificial neural networks for habitat suitability modelling). Both approaches have strengths and weaknesses and have a proper context for its application. The idea with this research is to combine the strengths of both approaches by applying each of them in the proper context for predicting macroinvertebrates. This methodology can be seen as an integrated way of ecological modelling, needed for the proper assessment of anthropogenic disturbances in aquatic ecosystems.

De Laender *et al*., (2005) presented an example of the combination of mechanistic and datadriven models in the framework of an ecological model for risk assessment, for estimating the effect on the ecosystem given a defined exposure concentration of the chemical of concern. ANN were used to predict habitat suitability in a river given some measured structural and physical variables (Figure 5.5). These results serve as an indication of the ecological status of the system and can subsequently be used as input for a mechanistic ecosystem model. Only those populations whose presence is predicted by the ANN are implemented in the mechanistic model. This allows for the simulation of the dynamics of these populations, given the environmental boundary conditions and chemical contamination due to anthropogenic disturbances.

5.3.2 Statistical modelling for various species and community data

When the aim of the research is the analysis of multiple species and community data, the direct gradient analysis methods are good statistical approaches. Canonical Correspondence Analysis (CCA) and Redundancy Analysis (RDA) are two direct gradient

analyses techniques that have been used in statistical habitat modelling. These methods are designed to detect the pattern of species abundance that can best be explained by observed environmental variables, and use a supplied matrix of predictor environmental variables to quantify the variation in a matrix of response variables. This property makes direct ordination methods a useful tool in aquatic habitat modelling of multiple species (Ahmadi-Nedushan *et al*, 2006). In RDA, species responses are assumed to be linear along an environmental gradient but CCA is based on nonlinear responses. This property makes CCA more appropriate in habitat modelling. In CCA, the species ordination is done directly and iteratively in relation to supplied environmental variables. CCA relates species abundance to measured variation in the environment by requiring the ordination axes to be linear combinations of the environmental variables (Ahmadi-Nedushan *et al*, 2006).

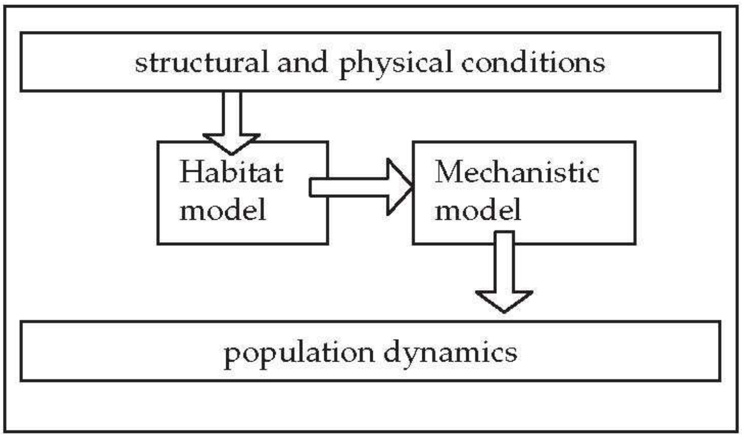


Figure 5.5 Flow chart of proposed methodology. Source: De Laender *et al*., (2005).

5.3.3 Hydromorphological pressures and their effects on the biota

Hydromorphological pressures refers to all changes caused by human influences to either the flow regime (hydro) or the morphology of the stream that affect the stream biota. The most important hydromorphological pressures are**:** building of dams or weirs for hydropower, water supply or other purposes; channelisation and/or dredging of rivers or streams to improve drainage or for navigation; weed cutting to improve drainage; abstraction of water directly from the stream or from groundwater for water supply or irrigation, or diversion (REBECCA, 2004).

For the Cauca river, it is important to consider the impact of the hourly river flow fluctuations generated by the operation of the Salvajina dam, which affects the dilution capacity of the river and can affect benthos communities, with a reduction in species richness, while some species increase in abundance (Fruget, 1991). With this study macroinvertebrate habitat evaluations will be included as a component of minimum flow evaluations, for release schedules for the hydropower facility Salvajina. It is important as well to study the hourly impact of the release schedules of the hydropower facility Salvajina on the dissolved oxygen (DO) in the Cauca river, because until now the evaluation of the scenarios for pollution control proposed by the environmental authority CVC has considered as a result an increase in the average values of DO in the river, however the dynamics of the system makes necessary an evaluation of the percentage of the time that the DO is below certain reference value. Prolonged exposure to low DO levels (<5 mg/l) may not directly kill an organism, but will increase its susceptibility to other environmental stresses. Exposure to DO < 2 mg/l for one to four days may kill most of the biota in a system.

6. CONCLUSIONS

6.1 GENERAL CONCLUSIONS

This thesis has demonstrated the high potential of the integration of hydrodynamic and physical-chemical water quality models with ecological models, in helping to get insight in aquatic ecosystems, for finding what is necessary to improve in the integrated water management and policy development.

The integrated ecological model proposed in this research is a powerful operational tool, which allows to model and to assess the ecological impact of wastewater discharges into the Cauca river and can help to calculate the needed reductions in wastewater discharges of organic matter to meet biological quality criteria in this river.

Today river 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 are complementary tools for river quality assessment. 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.

6.2 CONCLUSIONS OF STATISTICAL PREDICTIVE MODELS FOR MACROINVERTEBRATES

The statistical models proposed in this research, allow predicting the occurrence and the abundance of macroinvertebrates for the Cauca river under different hydraulic and physical-chemical water quality conditions.

The assessment of the multiple logistic regression models (MLRMs) reliability showed that the models for Ephemeroptera (AUC=1), Trichoptera (AUC=1), and Haplotaxida (AUC=0.926) correctly discriminates between occupied (presence) and unoccupied (absence) sites in the dataset. Regarding the predictive validation procedure for quasi-Poisson regression models (QPRMs), it was found that in general the models reproduce with good precision the tendencies and the maximum and minimum values of abundance data for each macroinvertebrate (i.e. Ephemeroptera, Trichoptera and Haplotaxida) and the BMWP index, with high R^2 values ($0.866 < R^2 < 0.998$). However, it was found that the QPRM for Haplotaxida model performs better at high abundance values rather than low abundance values.

6.3 CONCLUSIONS OF THE INTEGRATED ECOLOGICAL MODELLING OF THE CAUCA RIVER

The application of the integrated ecological modelling of the Cauca river showed that the MLRMs and QPRMs predict well the ecological impact of the scenarios for pollution control in the Cauca river's basin. Thus, in the scenario with the highest pollution reduction (very optimistic scenario, E8-year 2015) an improvement of the water quality of the Cauca river is achieved, which is represented with the presence and/or an increase of the number of pollution sensitive benthos (i.e. Ephemeroptera and Trichoptera) and the absence and/or a decrease of the number of pollution tolerant benthos (i.e. Haplotaxida). On the other hand, if the worst pollution condition scenario is considered (very pessimistic scenario, E12-year 2015) a deterioration of the water quality is obtained, which is represented with the absence and/or a decrease of the number of pollution sensitive benthos and the presence and/or an increase of the number of pollution tolerant benthos. Additionally, the BMWP index predictions for these scenarios confirm the previous results, showing an increase of the BMWP values in some monitoring stations in the case of the very optimist scenario (E8) indicating water quality improvement and a reduction of the BMWP values in the case of the very pessimistic scenario (E12) indicating water quality deterioration.

REFERENCES

Addinsoft. 2009. 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.

Adriaenssens V., Goethals P.L.M., De Pauw N. 2006. Fuzzy knowledge-based models for prediction of Asellus and Gammarus in watercourses in Flanders (Belgium). Ecological Modelling. Vol 195. Pags 3–10.

Adriaenssens, V., Goethals P.L.M., Charles, J., De Pauw N. 2004. Application of Bayesian belief networks for the prediction of macroinvertebrate taxa in rivers. Annales de Limnologie – International Journal of Limnology, 40(3), 181-191.

Ahmadi-Nedushan B., ***et al*****. 2006.** A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment. River Research and Applications 22:503-523.

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

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

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

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

CETESB. 2002. Technology company for the environmental sanitation of Brasil. www.cetesb.sp.gov.br/agua/rios/indice.asp. Water quality index for drinking water use.

Chapman S., *et al.* 1996. Water Quality Assessments: A Guide to the Use of Biota, Sediments, and Water in Environmental Monitoring. Second edition. Spon Press. London England.

Chapra, S. C. 1997. Surface Water-quality Modeling. New York : McGraw-Hill.

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., Steven.Chapra@tufts.edu

CVC-Corporación Autónoma Regional del Valle del Cauca. 2005. Pollutant organic load discharges in the Cauca's river basin reported for the year 2005. Santiago de Cali, Colombia. Available in Spanish.

CVC-Corporación Autónoma Regional del Valle del Cauca, Univalle-Universidad del Valle. 2007a. Modelling scenarios for defining pollution control plans in the Cauca's river basin. Streach La Balsa – Anacaro. Volume XII. Cauca River Modelling Project (PMC), Phase III. Santiago de Cali, Colombia. Available in Spanish.

CVC-Corporación Autónoma Regional del Valle del Cauca, Univalle-Universidad del Valle. 2007b. Optimization of the water quality simulation model for the Cauca river. Streach La Balsa – Anacaro. Volume XIII. Cauca River Modelling Project (PMC), Phase III. Santiago de Cali, Colombia. Available in Spanish.

CVC-Corporación Autónoma Regional del Valle del Cauca, Univalle-Universidad del Valle. 2007c. Water quality of the Cauca River and its tributaries, Chapter 7. and Mathematical Modelling of the Cauca river, Chapter 8. The Cauca river at its high Valley: a contribution to the knowledge of one of the most important Colombian rivers. Colombia. pages: 207-266. Available in Spanish.

CVC-Corporación Autónoma Regional del Valle del Cauca, Univalle-Universidad del Valle. 2004a. Identification of critical parameters in the Cauca river and its main tributaries. Volume IX. Cauca River Modelling Project (PMC), Phase II. Santiago de Cali, Colombia. Available in Spanish.

CVC-Corporación Autónoma Regional del Valle del Cauca, Univalle-Universidad del Valle. 2004b. Water quality study of the Cauca river and its main tributaries by means of the application of water quality and pollution indices. Stretch Salvajina – La Virginia. Volume X. Cauca River Modelling Project (PMC), Phase II. Santiago de Cali, Colombia. Available in Spanish.

CVC-Corporación Autónoma Regional del Valle del Cauca, Univalle-Universidad del Valle. 2004c. Hydrobiology (benthic macroinvertebrates, phytoplankton and zooplankton) and water quality assessment throughout nine (9) monitoring stations in the Cauca river. Volume XII. Cauca River Modelling Project (PMC), Phase II. Santiago de Cali, Colombia. Available in Spanish.

CVC-Corporación Autónoma Regional del Valle del Cauca, Univalle-Universidad del Valle. 2004d. Calibration and applications of the water quality model for the Cauca river. Volume VIII. Cauca River Modelling Project (PMC), Phase II. Santiago de Cali, Colombia. Available in Spanish.

De Laender F., Goethals P.L.M., Vanrolleghem P., Janssen C. 2005. Ecological effect assessment through the combination of mechanistic and data driven models. Proceedings COST 626–European Aquatic Modelling Network, final meeting in Silkeborg, Denmark 19-20 May 2005. (In Harby, A. *et al* (editors) 2005).

De Laender Frederik. 2007. Predicting effects of chemicals on freshwater ecosystems model development, validation and application. Doctoral Thesis in Applied Biological Sciences. Ghent University. Gent, Belgium.

De Pauw N., Gabriels W., Goethals P.L.M. 2006. River Monitoring and Assessment Methods Based on Macroinvertebrates. Chapter 7. In the book Biological Monitoring of Rivers. Editor(s): Ziglio G, Siligardi M., Flaim G. Series: Water Quality Measurements Series. John Wiley & Sons, Ltd.

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.

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.

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

DHI-Danish Hidraulic Institute. 1999. MIKE11, A modelling system for rivers and channels, Reference manual, DHI Water & Environment, Hørsholm, Denmark.

Dolan David M., El-Shaarawi Abdel, Reynoldson Trefor B. 2000. Predicting benthic counts in Lake Huron using spatial statistics and quasi-likelihood. Environmetrics Vol 11. Issue 3. Pages 287-304.

Domínguez L. 2007. Indices based on macroinvertebrate communities for assessment of the quality of the Chaguana river in Ecuador. Doctoral Thesis in Applied Biological Sciences. Ghent University. Gent, Belgium.

Domínguez L., Goethals P.L.M., De Pauw N. 2005 Development of a biological assessment methodology based on macroinvertebrates in the Chaguana river basin (Ecuador). En: Seminario Internacional La Hidroinformática en la Gestión Integrada de los Recursos Hídricos. AGUA 2005. Santiago de Cali, Colombia.

Dunnette, D.A. 1979. A Geographically Variable Water Quality Index Used in Oregon. Journal of the Water Pollution Control Federation 51(1): 53-61

EMCALI-Empresas Municipales de Cali, Univalle-Universidad del Valle. 2006. Impact assessment of the proposed strategies by EMCALI for the management of wastewaters in the city of Cali on the water quality of the Cauca river. Available in Spanish.

EMCALI-Empresas Municipales de Cali. 2007. Plan for the sanitation and the wastewater discharge management, 2007-2016, Cali, Colombia. Available in Spanish.

EU-European Union. 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

Fernandez, N., and Solano F. 2008. Water Quality Indices (WQIs) and Water Pollution Indices. Chapter IV. A comparative analyses of Water Quality Indexes (WQIs) and Water Pollution Indices indexes. Pamplona University. Pamplona, Colombia. Available in Spanish.

Fruget J.F. 1991. The impact of river regulation on the lotic macroinvertebrate communities of the lower Rhone, France. Regulated Rivers: Research & Management 6: 241-255.

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

Galvis A. 2004. Course notes of Models in Sanitary and Environmental Engineering. Academic Year 2003-2004. Universidad del Valle - Facultad de Ingeniería. Instituto Cinara.Cali, Colombia.

Giller Paul S. and Malmqvist Björn. 1998. The biology of streams and rivers, Oxford University Press.296 pages

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. 2007. Course notes Environmental Biology. Academic Year 2007-2008. Faculty of Bioscience Engineering, Ghent University, Belgium.

Goethals P.L.M., Dedecker A., Gabriels W., De Pauw N. 2006. Development and application of predictive river ecosystem models based on classification trees and artificial neural networks. Chapter 8. Book: Ecological informatics: Scope, Techniques and applications, 2nd Edition. Recknagel, F. (ed.), Springer Berlin Heidelberg.

Gordon Johnston. 2008. SAS Software to fit the Generalized Linear Model. SAS Institute Inc., Cary, NC.

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.

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. Vol 157. Pages 89- 100.

Gutiérrez J., Riss W., Ospina R. 2004a. Bioindication of the water quality with aquatic macroinvertebrates in the sabanna of Bogota using artificial neural network. In: Journal Caldasia. Limnologia. Vol 26 (1). Pages 151-160. Available at: www.unal.edu.co/icn/publicaciones/caldasia.htm

Gutiérrez J., Riss W., Ospina R. 2004b. Application of fuzzy logic as bioindication tool for the water quality with aquatic macroinvertebrates in the sabanna of Bogota-Colombia. In: Journal Caldasia. Limnologia. Vol 26(1). Pages 161-172. Available at: www.unal.edu.co/icn/publicaciones/caldasia.htm

Hastie, T.J., Tibshirani, R. 1990. Generalized Additive Models. Chapman and Hall, London.

Hosmer David and Lemeshow Stanley. 2000. Applied Logistic Regression. United States of America, Wiley-Interscience Publication. 373 pages

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

INGESAM. 2005. Environmental impact study of the Navarro's dump in the area of deferred regimen of Navarro and surrounding districts, called: Melendez, City 2000, Cordoba city, The Caney, COMFANDI citadel, Juanambú and Encuentros colleges and Cañas gordas Club. Report of technical characterization of the Navarro's dump. Available in Spanish.

Irvine K. *et al.* 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.

Jones M.T., Niemi G.J., Hanowski J.M., Regal R.R. 2002. Poisson regression: a better approach to modeling abundance data. In: Scott, J.M., Heglund, P.J., Morrison M. *et al*, Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Covelo, CA.

Jorgensen S.E. 2008. Overview of the model types available for development of ecological models. Ecological modeling. Vol 215. Pages 3–9

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

Lek Sovan. 2007. Uncertainty in ecological models. Editorial. Ecological modeling. Volume 207. Pages 1-2.

Lindsey James. K. 1997. Applying Generalized Linear Models. Springer. 256 pages

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.

Mackenthun, K.M. 1969. The practice of water pollution biology. FWPCA, Washington, pp. 281.

Manel S., Buckton S.T., Rormerod S.J. 2000. Testing large-scale hypotheses using surveys: the efects of land use on the habitats, invertebrates and birds of Himalayan rivers. Journal of Applied Ecology 37:756-770.

Manel S., Dias J-M., Steve O. J. 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. Ecological Modelling. Volume 120. Pages 337-347.

Martin J., McCutcheon S., Schottman R. 1999. Hydrodynamics and transport for water quality modeling. Edition: illustrated. published by CRC Press, 794 pages.

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

Milner A.M., Brittain J.E., Castella E., Petts G.E. 2001. Trends of macroinvertebrate community structure in glacier-fed rivers in relation to environmental conditions: a synthesis. Freshwater Biology. Vol 46. Pages 1833-1847.

Mouton A. M., De Baets B., Goethals P.L.M. 2009a. Knowledge-based versus data-driven fuzzy habitat suitability models for river management." Environmental Modelling & Software 24(8): 982-993.

Mouton A.; Van Der Most H., Jeuken A; Goethals P.L.M.; De Pauw N. 2009b. Evaluation of river basin restoration options by the application of the Water Framework Directive Explorer in the Zwalm river basin (Flanders, Belgium). In: River Research and Applications. Volume 25 Issue 1, Pages 82 – 97

Nestler J., Goodwin R. A., Loucks D. P. 2005. Coupling of Engineering and Biological Models for Ecosystem Analysis. In: Journal of Water Resources Planning and Management, Vol. 131, No. 2, March 1. Pages 101 – 109

NIVA-Norwegian Institute for Water Research. 2007. Statistical and modelling methods for assessing the relationships between ecological and chemical status in different lake types and different geographical regions. Deliverable 12. Final Conference, REBECCA project (Relationships between ecological and chemical status of surface waters). Edited by S. Jannicke Moe and Robert Ptacnik. Oslo, Norway

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

Piñeiro Gervasio, Perelmanb Susana, *et al*. 2008. How to evaluate models. Observed vs. predicted or predicted vs. observed. Ecological modelling. Vol 2 1 6. Pages 316–322.

Ramírez A., Restrepo R., Cardeñosa M. 1999. Water pollution indices for characterization of continental waters and discharges. Formulations. Science, Technology and future. Vol 1. Issue 5. Pages 89-99.

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

Ramirez A., Viña V. G. 1998. Water quality criteria and its relation with the benthos in the influence area of the oil pipeline Cusiana-Coveñas. Memories water quality environmental bioindicators.

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. *Wat. Sci. Tech.*, **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

Reineking Bjorn, Schroderc Boris. 2006. Constrain to perform: Regularization of habitat models. Ecological Modelling. Vol 193. pages 675-690.

Rodrigues M., Oliveira A., Queiroga H., Zhang Y. J., Fortunato B., Baptista M. 2007. Integrating a Circulation Model and an Ecological Model to Simulate the Dynamics of Zooplankton. Estuarine and Coastal Modeling Congress 2007. Part of Estuarine and Coastal Modeling Congress 2007. Proceeding of the Tenth International Conference on Estuarine and Coastal Modeling Congress 2007 Newport, Rhode Island, USA.

Roldan G. 1992. Fundamentals in tropical limnology. Editorial Universidad de Antioquía, Medellín, Colombia. 529 p.

Rushton S.P., Ormerod S.J., *et al*. 2004. New paradigms for modelling species distributions? Journal of Applied Ecology 41(2): 193-200.

Saito L. 1999. Interdisciplinary modelling at Shasta Lake. Fort Collins: Colorado State University. 341 p. Ph.D. dissertation.

Saito L., Segale H. 2006. Interdisciplinary Modelling of Aquatic Ecosystems Summary report Curriculum Development Workshop. California

Sudhir Paul and Krishna K. Saha. 2007. The generalized linear model and extensions: a review and some biological and environmental applications. Environmetrics. Vol 18. Issue 4. pages 421-443.

UN-United Nations. 1992. The Dublin Statement on Water and Sustainable Development. International Conference on Water and the Environment0: Development issues for the 21st century. 26-31 January 1992, Dublin, Ireland.

UN-United Nations. 2003. Water for People, Water for Life. UN World Water Development Report. Prepared as a collaborative effort of 23 UN agencies and convention secretariats co-ordinated by the World Water Assessment Programme. UNESCO, Paris. http://www.unesco.org/water/wwap/index.shtml

Velez Carlos A. 2006. Integrated Water Quality and Ecosystem Modelling a Case Study for Sonso Lagoon, Colombia. Master of Science thesis. UNESCO-IHE Institute for Water Education, Delft, The Netherlands.

Ver Hoef J. M., Boveng, P. L. 2007. Quasi-Poisson vs. Negative Binomial Regression: how should we model overdispersed count data? Ecology Volume 88 (11). Pages 2766-2772.

White Gary C., Bennetts Robert E. 1996. Analysis of Frequency Count Data Using the Negative Binomial Distribution. Ecology 77(8): 2549-2557.

Williams Brian. 2006. Hydrobiological Modelling: Processes, Numerical Methods and Applications. Published by Lulu .com. 700 pages.

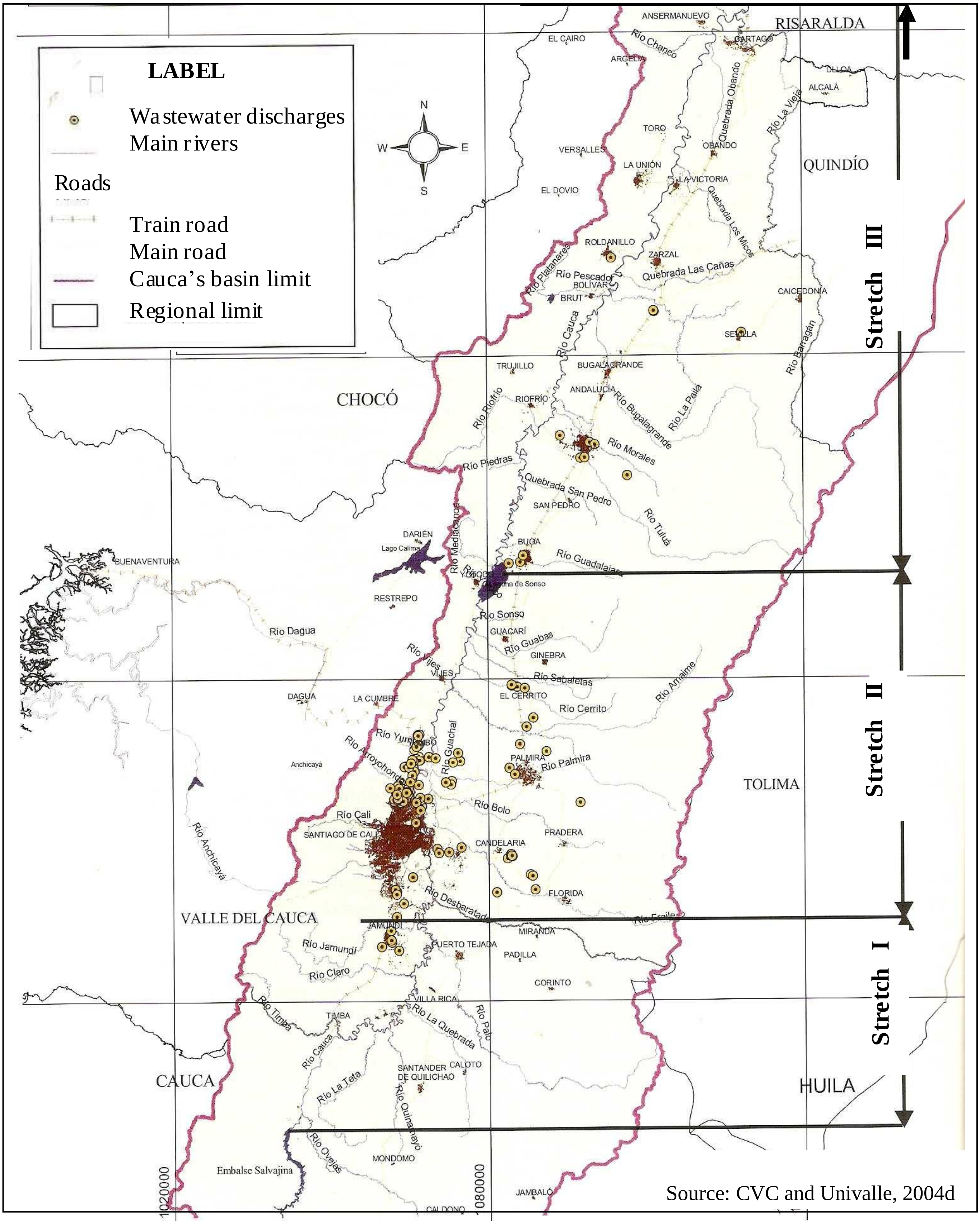
Zúñiga M. del C. 2009. Chapter 7. Water quality and environmental flow bioindicators. In: Cantera, J, Y. Carvajal y L. Castro. Environmental flow: Concepts, experiences and challenges concepts. Programa Editorial de la Universidad del Valle, Cali, Colombia.

Web sources:

ODEQ. 1980. Oregon's Department of Environmental Quality. Oregon Water Quality Index. http://www.deq.state.or.us/lab/wqm/wqindex.htm

Annexes

Annex A. Location of domestic and industrial wastewater discharges which are monitored by the CVC in the Cauca river's basin



Annex B. Database selected for the implementation of the ecological models for predicting macroinvertebrates in the Cauca river

Table B1. Physical-Chemical and Hydraulic database. Cauca River. 1997 - 2004

Sampling site	Date	Condition	Flow (m3/s)	Depth (m)	Velocity (m/s)	Substrate % Sand	% Gravel	% Silt & clay	Temp (°C)	DBO_5 (mg/l)	DO (mg/l)	pH (unidad)	N Total (mg/l)	$N\text{-}NH_4$ (mg/l)	$N\text{-}NO_2$ (mg/l)	$N\text{-}NO_3$ (mg/l)
Paso de la Balsa	19-Feb-97	High flow	269	4.9	1.2	28.50	69.86	1.62	18.0	3.7	6.8	6.4	2.010	*	0.0026	0.1906
Paso de la Bolsa	19-Feb-97	High flow	311	4.2	0.96	93.20	6.42	0.43	20.0	1.9	5.9	6.7	*	*	0.0032	0.2370
Juanchito	19-Feb-97	High flow	395	7.2	0.9	0.99	0.52	0.34	20.0	3.2	5.6	6.6	1.500	0.000	*	0.345
Paso de la Torre	19-Feb-97	High flow	419	6.9	0.9	0.89	0.11	0.06	20.0	4.6	4.5	6.7	1.880	0.000	0.0117	0.2745
Mediacanoa	19-Feb-97	High flow	509	6.3	0.9	0.93	7.17	0.07	22.0	6.8	3	6.7	2.360	0.000	0.0192	0.3299
Anacaro	25-Jul-01	Low flow	177	3.3	0.52	97.60	2.24	0.16	26.4	2.5	2.62	7.3	*	1.700	0.0660	0.8540
Paso de la Balsa	22-Jun-04	Low flow	83	3.4	0.62	28.50	69.86	1.64	24.0	0.23	6.89	7.07	1.590	*	*	0.2170
Paso de la Bolsa	22-Jun-04	Low flow	109	2.8	0.55	93.20	6.40	0.40	22.0	0.12	6.67	7.06	3.240	*	*	0.3090
Puente Hormiguero	22-Jun-04	Low flow	120	2.1	0.57	77.90	22.00	0.10	23.0	0.49	5.94	7.06	1.800	*	*	0.4590
Antes Navarro	23-Jun-04	Low flow	128	5.0	0.58	77.90	22.00	0.10	25.0	1.44	5.06	7.37	0.230	*	*	0.2860
Juanchito	23-Jun-04	Low flow	132	4.9	0.58	99.10	0.54	0.36	23.0	1.68	4.56	8.17	1.480	*	*	0.357
Paso de la Torre	23-Jun-04	Low flow	123	4.8	0.56	88.50	11.45	0.05	24.0	15.45	0.30	7.37	2.930	*	*	0.3020
Mediacanoa	23-Jun-04	Low flow	152	3.7	0.55	92.80	7.14	0.06	23.0	12.75	0.82	7.0	1.180	*	*	0.2620
Puente la Victoria	24-Jun-04	Low flow	165	2.4	0.55	97.60	2.24	0.16	24.0	3.3	2.28	7.34	1.18	*	*	0.555
Anacaro	24-Jun-04	Low flow	187	3.4	0.54	97.60	2.24	0.16	25.5	3.38	1.58	7.29	1.870	*	*	0.4700

* Parameters do not included during the physical-chemical and hydrobiological monitorings

Source: CVC and Univalle, 2004b

Annex B. Database selected for the implementation of the ecological models for predicting macroinvertebrates in the Cauca river

Table B1. Physical-Chemical and Hydraulic database. Cauca River. 1997 – 2004 (cont.)

Sampling site	Date	P Total (mg/l)	P-PO4 (mg/l)	Colour (UPC)	Turbidity (FTU)	Total solids (mg/l)	Suspended Solids (mg/l)	Dissolved Solids (mg/l)	COD (mg/l)	Total Hardness (mg/l)	Ca Hardness (mg/l)	Mg Hardness (mg/l)	Calcium (mg/l)	Magnesium (mg/l)
Paso de la Balsa	19-Feb-97	0.090	0.038	75	56.7	139	76	63	20.6	20	10.5	9.5	4.2	2.3
Paso de la Bolsa	19-Feb-97	0.064	0.036	75	68	154	101.4	52.6	20.6	21.4	11	10.4	4.4	2.5
Juanchito	19-Feb-97	0.204	0.046	150	190	336	254.3	81.7	37.2	33.3	19.8	13.5	7.9	3.2
Paso de la Torre	19-Feb-97	0.212	0.046	75	105	259	182	76.2	29.9	39.7	23	16.7	9.2	4
Mediacanoa	19-Feb-97	0.243	0.029	50	125	349	242	106.2	37.2	43.9	24.7	19.2	9.9	4.6
Anacaro	25-Jul-01	0.072	0.060	40	91	185	71	114	17.30	60	38	22	15.2	5.28
Paso de la Balsa	22-Jun-04	< 0.03	0.015	14.8	13	68	14.3	54	*	30	16	14	6.4	3.36
Paso de la Bolsa	22-Jun-04	< 0.03	0.017	18.2	14	88	17	71	*	27	18	9	7.2	2.16
Puente Hormiguero	22-Jun-04	0.067	0.036	23.5	50	103	30.2	73	*	35	20	15	8	3.6
Antes Navarro	23-Jun-04	0.061	0.054	25.6	25	112	28.2	84	*	38	28	10	11.2	2.4
Juanchito	23-Jun-04	0.099	0.077	21.3	29	98	24	74	*	40	22	18	8.8	4.32
Paso de la Torre	23-Jun-04	0.360	0.190	24	12	119	28	91	*	47	30	17	12	4.08
Mediacanoa	23-Jun-04	0.233	0.111	23	18	158	54	105	*	57	33	24	13.2	5.76
Puente la Victoria	24-Jun-04	0.279	0.037	16	64	170	73.7	96	*	80	42	38	16.8	9.12
Anacaro	24-Jun-04	0.176	0.053	11	55	157	58	99	*	58	35	23	14	5.52

* Parameters do not included during the physical-chemica l and hydrobiological monitorings
Source: CVC and Univalle, 2004b

Annex B. Database selected for the implementation of the ecological models for predicting macroinvertebrates in the Cauca river

Table B1. Physical-Chemical and Hydraulic database. Cauca River. 1997 – 2004 (cont.)

Sampling site	**Date**	**Total Alkalinity (mg/l)**	**Bicarbonate (mg/l)**	**Conductivity (uS/cm)**	**Fe Total (mg/l)**	**Mn Total (mg/l)**	**Na Total (mg/l)**	**K Total (mg/l)**	**Cu Total (mg/l)**	**Zn Total (mg/l)**	**Chloride (mg/l)**	**SO_4^{-2} (mg/l)**	**Total Coliforms (NMP/100 ml)**	**Faecal Coliforms (NMP/100 ml)**
Paso de la Balsa	19-Feb-97	14	17.1	62.6	8.12	0.150	3.830	1.470	0.002	0.082	3.310	14.680	2.4E+03	2.4E+02
Paso de la Bolsa	19-Feb-97	16	19.5	63.2	12.12	0.230	5.540	1.680	0.002	0.086	2.730	17.600	9.4E+03	2.4E+02
Juanchito	19-Feb-97	24	29.3	84.7	13.30	0.260	3.510	1.680	0.002	0.160	2.680	20.170	2.4E+08	2.4E+02
Paso de la Torre	19-Feb-97	35	42.7	113.4	8.35	0.200	4.670	1.650	0.002	0.065	3.850	21.870	2.4E+06	2.4E+03
Mediacanoa	19-Feb-97	42	51.2	120	14.48	0.340	5.420	1.760	0.002	0.106	4.280	20.910	2.4E+06	2.4E+03
Anacaro	25-Jul-01	50.8	61.9	165	3.73	0.133	9.200	1.970	0.020	0.030	7.820	19.900	2.4E+07	2.4E+06
Paso de la Balsa	22-Jun-04	22.9	27.9	75.3	*	*	*	*	*	*	6.440	*	2.40E+ 0.4	2.40E + 0.3
Paso de la Bolsa	22-Jun-04	18.6	22.7	76.9	*	*	*	*	*	*	7.320	*	2.40E+ 0.3	2.40E+ 0.3
Puente Hormiguero	22-Jun-04	26.4	32.2	94.1	*	*	*	*	*	*	7.710	*	2.4E+0.5	2.4E+0.5
Antes Navarro	23-Jun-04	32.6	39.8	104.6	*	*	*	*	*	*	7.280	*	2.4E+0.5	2.4E+0.5
Juanchito	23-Jun-04	32.7	39.9	130.7	*	*	*	*	*	*	7.220	*	2.4E+0.4	2.4E+0.4
Paso de la Torre	23-Jun-04	55.3	67.4	759	*	*	*	*	*	*	10.330	*	2.4E+0.6	2.40E+0.6
Mediacanoa	23-Jun-04	44.4	54.1	152	*	*	*	*	*	*	8.740	*	2.4E+0.3	2.4E+0.3
Puente la Victoria	24-Jun-04	54.7	66.7	159.5	*	*	*	*	*	*	9.87	*	2.40E+0. 4	2.40E+0.4
Anacaro	24-Jun-04	54.4	66.4	160	*	*	*	*	*	*	9.570	*	2.4E+0.4	2.4E+0.4

* Parameters do not included during the physical-chemical and hydrobiological monitorings
Source: CVC and Univalle, 2004b

Annex B. Database selected for the implementation of the ecological models for predicting macroinvertebrates in the Cauca river.(cont.)

Table B2. Hidrobiological database (Benthic macroinvertebrates). Cauca River. 1997

Station	Phyllum	Class	Order	Family	Genus	N	Relative frecuency (%)
Station Paso de La Balsa Febrero	Arthropoda	Insecta	Ephemeroptera	Baetidae	*Baetis sp.*	13	26.0
				Leptophlebiidae	*Thraulodes sp.*	14	28.0
			Odonata	Aesnidae	*Aeshna sp.*	5	10.0
				Libellulidae	*Dythemis sp.*	2	4.0
			Trichoptera	Hydropsychidae	*Smicridea sp.*	7	14.0
					Leptonema sp.	3	6.0
			Neuroptera	Corydalidae	*Corydalus sp.*	6	12.0
	TOTAL					50	100.0
Station Paso de La Bolsa Febrero	Arthropoda	Insecta	Trichoptera	Leptoceridae	*Atanatolica sp.*	23	31.1
			Odonata	Libellulidae	*S.I*	11	14.9
			Coleoptera	Elmidae	*Macrelmis sp.*	7	9.5
			Diptera	Simuliidae	*Simulium sp.*	18	24.3
				Tipulidae	*Hexatoma sp.*	3	4.1
			Neuroptera	Corydalidae	*Corydalus sp.*	12	16.2
	TOTAL					74	100.0
Station Juanchito Febrero	Arthropoda	Insecta	Diptera	Tipulidae	*Hexatoma sp.*	22	30.1
				Simuliidae	*Simulium sp.*	16	21.9
				Blepharoceridae	*Limonicola sp.*	7	9.6
			Trichoptera	Hydropsychidae	*Leptonema sp.*	23	31.5
			Odonata	Libellulidae	*Brechmorhoga sp.*	5	6.8
	TOTAL					73	100.0
Station Paso de la Torre Febrero	Arthropoda	Insecta	Diptera	Syrphidae	*Eristalis sp.*	22	13.5
	Annelida	Oligochaeta	Haplotaxida	Tubificidae	*Tubifex sp.*	53	32.5
	Mollusca	Gastropoda	Basommatophora	Physidae	*Physa sp.*	66	40.5
	Platyhelminthes	Turbellaria	Tricladida	Planariidae	*Dugesia sp.*	22	13.5
	TOTAL					163	100.0
Station Mediacanoa Febrero	Annelida	Oligochaeta	Haplotaxida	Tubificidae	*Tubifex sp.*	58	67.4
	Arthropoda	Insecta	Diptera	Simuliidae	*Simulium sp.*	22	25.6
			Coleoptera	Ptilodactylidae	*Anchytarsus sp.*	5	5.8
				Elmidae	*Cylloepus sp.*	1	1.2
	TOTAL					86	100.0

N: Number of collected organism

S.I: Specie without identification

Table B3. Hidrobiological database (Benthic macroinvertebrates). Cauca River. 2001

Station	Phyllum	Class	Order	Family	Genus	N	Relative frecuency (%)
Station Anacaro Junio	Arthropoda	Insecta	Trichoptera	Hydropsychiidae	*Leptonema sp.*	4	8.2
				Leptoceridae	*Nectopsyche sp.*	2	4.1
			Plecoptera	Perlidae	*Anacroneuria sp.*	13	26.5
			Odonata	Libellulidae	*Dythemis sp.*	6	12.2
					Hetaerina sp.	4	8.2
					Aeshna sp.	7	14.3
			Hemiptera	Gerridae	*Eugerris sp.*	6	12.2
			Coleoptera	Ptylodactylidae	*Anchitarsus sp.*	4	8.2
			Diptera	Blepharoceridae	*Limonicola sp.*	3	6.1
	TOTAL					49	100.0

N: Number of collected organism

S.I: Specie without identification

Annex B. Database selected for the implementation of the ecological models for predicting macroinvertebrates in the Cauca river. (cont.)

Table B4. Hidrobiological database (Benthic macroinvertebrates). Cauca River. 2004

Station	Phyllum	Class	Order	Family	Genus	N	Relative frecuency (%)
Station Paso de la Balsa	Arthropoda	Insecta	Coleoptera	Psephenidae	*Psephenos sp.*	1	2.2
				Elmidae	*Macrelmis sp.*	1	2.2
				Staphylinidae	*S.I*	1	2.2
			Diptera	Blepharoceride	*Paltostoma sp.*	1	2.2
			Ephemeroptera	Leptohyphidae	*Leptohyphes sp.*	4	8.7
				Leptophlebiidae	*Thraulodes sp.*	5	10.9
			Hemiptera	Veliidae	*Rhagovelia sp.*	24	52.2
			Trichoptera	Leptoceridae	*Atanatolica sp.*	4	8.7
					Grumichella sp.	1	2.2
				Hydropsychidae	*Smicridea sp.*	2	4.3
	Mollusca	Gastropoda	MesoGastropoda	Thiaridae	*Hemisinus sp.*	2	4.3
	TOTAL					46	100.0
Station Paso de la Bolsa	Arthropoda	Insecta	Coleoptera	Elmidae	*Heterelmis sp.*	1	2.8
			Ephemeroptera	Baetidae	*Camelobaetidius sp.*	5	13.9
				Leptohyphidae	*Leptohyphes sp.*	6	16.7
				Leptophlebiidae	*Thraulodes sp.*	1	2.8
			Hemiptera	Naucoridae	*Cryphocricos sp.*	1	2.8
				Veliidae	*Rhagovelia sp.*	16	44.4
			Odonata	Libellulidae	*Perithemis sp.*	1	2.8
			Trichoptera	Hydropsychidae	*Smicridea sp.*	1	2.8
				Leptoceridae	*Nectopsyche sp.*	2	5.6
	Mollusca	Gastropoda	MesoGastropoda	Thiaridae	*Hemisinus sp.*	2	5.6
	TOTAL					36	100.0
Station Hormiguero	Arthropoda	Insecta	Coleoptera	Staphylinidae	*S.I*	1	0.2
				Elmidae	*Macrelmis sp.*	1	0.2
			Ephemeroptera	Baetidae	*Camelobaetidius sp.*	3	0.5
				Leptohyphidae	*Leptohyphes sp*	1	0.2
			Trichoptera	Leptoceridae	Oecetis sp.	1	0.2
			Hemiptera	Naucoridae	*Limnocoris sp.*	1	0.2
	Annelida	Oligochaeta	Haplotaxida	Tubificidae	*Tubifex sp*	500	82.4
	Mollusca	Gastropoda	MesoGastropoda	Thiaridae	*Hemisinus sp.*	97	16.0
	Plathyhelminthes	Turbellaria	Tricladida	Planariidae	*S.I*	2	0.3
	TOTAL					607	100.0
Station Antes Navarro	Arthropoda	Insecta	Coleoptera	Elmidae	*Heterelmis sp.*	2	0.2
			Diptera	Chironomidae	*S.I*	164	15.7
			Ephemeroptera	Baetidae	*Camelobaetidius sp.*	2	0.2
				Leptohyphidae	*Leptohyphes sp.*	2	0.2
	Annelida	Oligochaeta	Haplotaxida	Tubificidae	*Tubifex sp.*	752	72.1
	Mollusca	Gastropoda	Basommatophora	Ancyliidae	*Ferrisia sp.*	41	3.9
			MesoGastropoda	Thiaridae	*Hemisinus sp.*	80	7.7
	TOTAL					1043	100.0
Station Juanchito	Arthropoda	Insecta	Diptera	Chironomidae	*S.I*	218	2.6
			Hemiptera	Veliidae	*Rhagovelia sp.*	10	0.1
	Annelida	Oligochaeta	Haplotaxida	S.I	*sp 1.*	1	0.0
				Tubificidae	*Tubifex sp.*	7500	89.2
				S.I	*sp2*	8	0.1
				S.I	*sp 3.*	1	0.0
	Mollusca	Gastropoda	MesoGastropoda	Thiaridae	*Hemisinus Sp.*	670	8.0
	TOTAL					8408	100.0
Station Paso de la Torre	Arthropoda	Insecta	Coleoptera	Hydrophilidae	*S.I*	1	0.0
			Diptera	Chironomidae	*S.I*	372	1.1
				Syrphidae	*Erystalis sp.*	1	0.0
	Annelida	Oligochaeta	Haplotaxida	S.I	*sp 1.*	1	0.0
				Tubificidae	*Tubifex sp.*	32500	98.7
				S.I	*sp 2.*	4	0.0
	Mollusca	Gastropoda	MesoGastropoda	Thiaridae	*Hemisinus sp.*	45	0.1
	TOTAL					32924	100.0

N: Number of collected organism

S.I: Specie without identification

Annex B. Database selected for the implementation of the ecological models for predicting macroinvertebrates in the Cauca river. (cont.)

Table B4. Hidrobiological database (Benthic macroinvertebrates). Cauca River. 2004. (cont.)

Station	Phyllum	Class	Order	Family	Genus	N	Relative frecuency (%)
Station Mediacanoa	Arthropoda	Insecta	Diptera	Chironomidae	*S.I*	65	1.1
			Odonata	Libellulidae	*Perithemis sp.*	1	0.0
	Annelida	Hirudinea	Glossiphoniiformes	Glossiphoniidae	*S.I*	2	0.0
		Oligochaeta	Haplotaxida	Tubificidae	*Tubifex sp.*	6000	97.8
				S.I	*sp. 2*	1	0.0
	Mollusca	Gastropoda	Basommatophora	Physidae	*Physa sp.*	1	0.0
			MesoGastropoda	Thiaridae	*Hemisinus sp.*	67	1.1
	TOTAL					6137	100.0
Station Puente la Victoria	Arthropoda	Insecta	Diptera	Chironomidae	*S.I*	56	4.8
				Blepharoceridae	*Limonicola sp.*	1	0.09
			Ephemeroptera	Leptohyphidae	*Leptohyphes*	3	0.3
			Hemiptera	Naucoridae	*Limnocoris sp.*	1	0.1
			Odonata	Gomphidae	*Perigomphus sp.*	1	0.1
					Aphilla sp.	1	0.1
	Annelida	Hirudinea	Glossiphoniiformes	Glossiphoniidae	*S.I*	13	1.1
		Oligochaeta	Haplotaxida	Tubificidae	*Tubifex sp.*	127	10.8
	Mollusca	Bivalbia	Unionoidea	Mycetopodidae	*Anodontites sp.*	3	0.3
		Gastropoda	MesoGastropoda	Thiaridae	*Hemisinus sp.*	970	82.5
	TOTAL					1176	100.0
Station Anacaro	Arthropoda	Insecta	Coleoptera	Elmidae	*Macrelmis sp.*	1	0.1
					S.I	12	0.7
			Hemiptera	Veliidae	*Rhagovelia sp.*	6	0.4
			Odonata	Gomphidae	*Phyllocycla sp.*	1	0.1
					Perigomphus sp.	1	0.1
	Annelida	Hirudinea	Glossiphoniiformes	Glossiphoniidae	*S.I*	2	0.1
		Oligochaeta	Haplotaxida	S.I	*sp. 1*	85	5.2
				Tubificidae	*Tubifex sp.*	110	6.8
	Mollusca	Gastropoda	Basommatophora	Ancylidae	*Ferrisia sp.*	45	2.8
			MesoGastropoda	Thiaridae	*Hemisinus sp.*	329	20.3
					S.I	1020	63.0
	Plathyhelminthes	Turbellaria	Tricladida	Planariidae	*S.I*	8	0.5
	TOTAL					1620	100.0

N: Number of collected organism

S.I: Specie without identification

Annex C. Results of physical-chemical and ecological indices used for assessing the Cauca's river water quality

Sampling site	Date	Condition	Physical-chemical water quality index		Ecological indices					
			ICAUCA		Abundance of species	Richness of species	Shannon Diversity		BMWP	
			Value	Water quality classification			Value	Classification	Value	Classification
Paso de la Balsa	19-Feb-97	High flow	64.1	Good	50	7	1.75	Half Polluted	43	Polluted
Paso de la Bolsa	19-Feb-97	High flow	*	*	74	6	1.63	Half Polluted	42	Polluted
Juanchito	19-Feb-97	High flow	50.1	Good	73	5	1.46	Quite Polluted	33	Quite Polluted
Paso de la Torre	19-Feb-97	High flow	43.2	Acceptable	163	4	1.27	Quite Polluted	9	Septic
Mediacanoa	19-Feb-97	High flow	34.7	Unsuitable	93	4	0.98	Quite Polluted	22	Quite Polluted
Anacaro	25-Jul-01	Low flow	*	*	49	9	2.05	Half Polluted	50	Polluted
Paso de la Balsa	22-Jun-04	Low flow	68.2	Good	46	11	1.69	Half Polluted	55	Polluted
Paso de la Bolsa	22-Jun-04	Low flow	67	Good	36	10	1.75	Half Polluted	45	Polluted
Puente Hormiguero	22-Jun-04	Low flow	45	Acceptable	607	9	0.74	Quite Polluted	29	Quite Polluted
Antes Navarro	23-Jun-04	Low flow	44.4	Acceptable	1043	7	0.88	Quite Polluted	22	Quite Polluted
Juanchito	23-Jun-04	Low flow	51.2	Good	8408	7	0.4	Quite Polluted	3	Septic
Paso de la Torre	23-Jun-04	Low flow	19.3	Extremely bad	32924	7	0.07	Quite Polluted	6	Septic
Mediacanoa	23-Jun-04	Low flow	31.5	Unsuitable	6137	7	0.13	Quite Polluted	16	Quite Polluted
Puente la Victoria	24-Jun-04	Low flow	37.6	Acceptable	1176	10	0.65	Quite Polluted	28	Quite Polluted
Anacaro	24-Jun-04	Low flow	35.1	Acceptable	1620	12	1.15	Quite Polluted	14	Quite Polluted

* Do not calculated because physical-chemical parameters for calculating the ICAUCA were not included during the monitorings

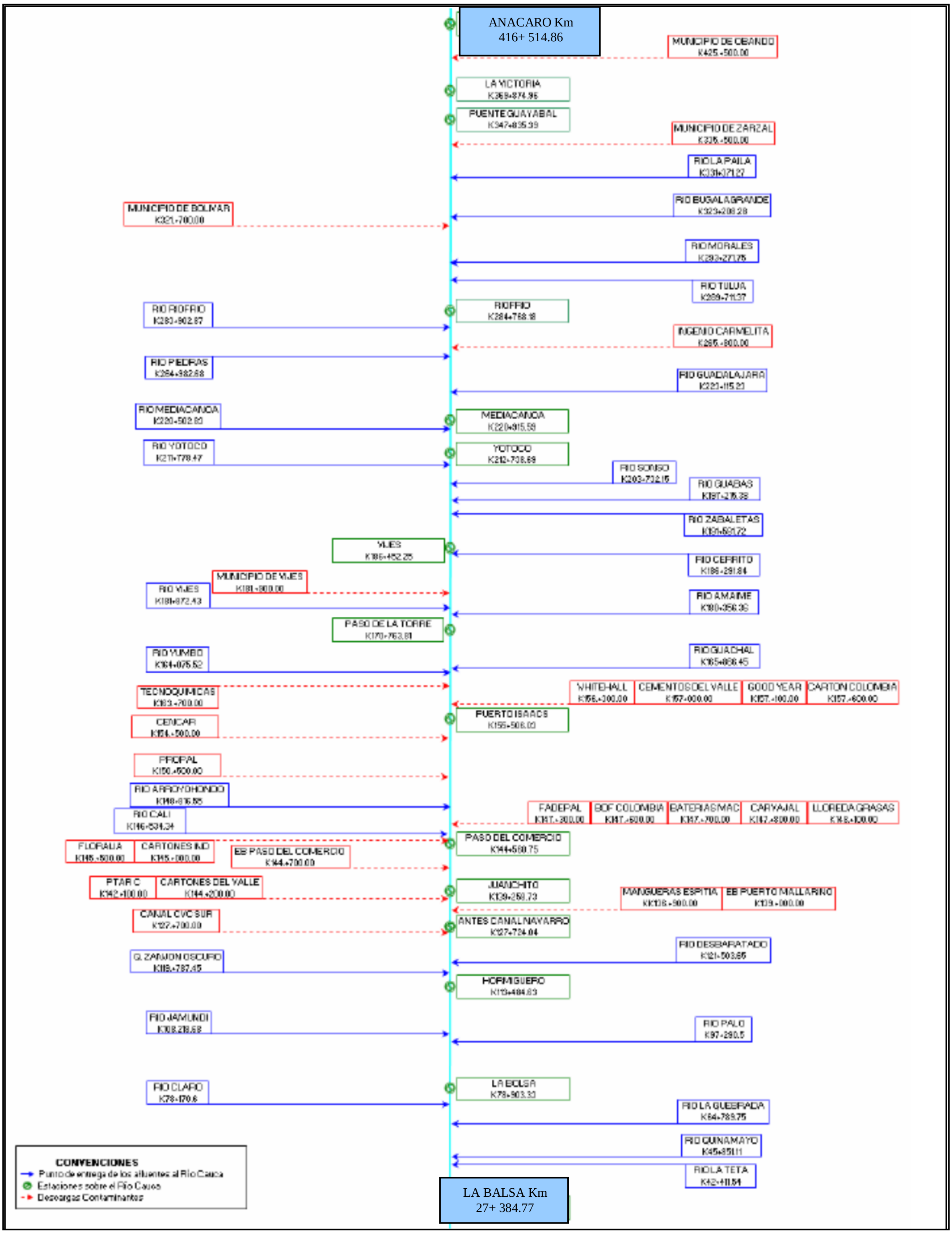
Source: CVC and Univalle, 2004b

Annex D. Scores assigned to each macroinvertebrate family according to their sensitivity to organic pollution used for the BMWP index calculation in the Cauca river.

Taxonomic Category		Score
Order	**Family**	
Plecoptera	Perlidae	10
Ephemeroptera	Oligoneuriidae	
Trichoptera	Calamoceratidae	
Coleoptera	Psephenidae	
Diptera	Blapharoceridae	
Odonata	Polythoridae	
Ephemeroptera	Euthyplociidae	9
Trichoptera	Helicopsychidae, Odontoceridae, Philopotamidae, Anomalopsychidae	
Coleoptera	Ptilodactylidae	
Megaloptera	Corydalidae	
Ephemeroptera	Leptophlebiidae, Polymitarcidae, Caenidae.	8
Trichoptera	Leptoceridae, Hidrobiosidae, Xiphocentronidae, Hydroptilidae	
Odonata	Gomphidae	
Ephemeroptera	Leptohyphidae	7
Trichoptera	Glossosomatidae, Polycentropodidae	
Coleoptera	Elmidae	
Odonata	Aeshnidae, Calopterygidae	
Coleoptera	Elmidae, Scyrtidae	6
Odonata	Coenagrionidae	
Diptera	Simuliidae	
Hemiptera	Corixidae, Gerridae, Veliidae	
Gasterópoda	Ancylidae	
Ephemerptera	Baetidae	5
Trichoptera	Hydropsychidae	
Coleoptera	Staphylinidae	
Odonata	Libellulidae	
Hemíptera	Naucoridae	
Diptera	Tipulidae, Simullidae	
Coleoptera	Curculionidae, Crysomelidae, Hydrophilidae, Gyrinidae, Diptera	4
	Tabanidae, Ceratopogonidae, Psychodidae, Dixidae, Empididae	
Lepidoptera	Pyralidae	
Hemiptera	Belostomatidae	
Tricladida	Planariidae	
Basommatophora	Lymneidae, Planorbidae	
Mesogastropoda	Thiaridae	
Coleoptera	Dyticidae	3
Hemiptera	Hydrometridae	
Basommatophora	Physidae,	
Bivalvia	Sphaeriidae	
Hirudinea	Glossiphonidae	
Diptera	Chironomidae, Culicidae, Syrphydae	2
Haplotaxida	Tubificidae	1

Source : Zúñiga (2009)

Annex E. Schematization of the Cauca river and tributaries



Source: CVC, Univalle, 2007b

Annex F. Database used in this research for developing predictive macroinvertebrate models

Table F.1 Macroinvertebrate abundance database

Sampling site	Date	Condition	Number of Organism collected							
			Coleóptera							
			Psephenidae	Elmidae (Macrelmis sp.)	Elmidae (Cylloepus sp.)	Elmidae (Heterelmis sp.)	Elmidae (Without. Id. sp)	Staphylinidae	Ptylodactylidae	Hydrophilidae
Paso de la Balsa	19-Feb-97	High flow	0	7	0	0	0	0	0	0
Paso de la Bolsa	19-Feb-97	High flow	0	0	0	0	0	0	0	0
Juanchito	19-Feb-97	High flow	0	0	0	0	0	0	0	0
Paso de la Torre	19-Feb-97	High flow	0	0	0	0	0	0	0	0
Mediacanoa	19-Feb-97	High flow	0	0	1	0	0	0	5	0
Anacaro	25-Jul-01	Low flow	0	0	0	0	0	0	4	0
Paso de la Balsa	22-Jun-04	Low flow	1	1	0	0	0	1	0	0
Paso de la Bolsa	22-Jun-04	Low flow	0	0	0	1	0	0	0	0
Puente Hormiguero	22-Jun-04	Low flow	0	1	0	0	0	1	0	0
Antes Navarro	23-Jun-04	Low flow	0	0	0	2	0	0	0	0
Juanchito	23-Jun-04	Low flow	0	0	0	0	0	0	0	0
Paso de la Torre	23-Jun-04	Low flow	0	0	0	0	0	0	0	1
Mediacanoa	23-Jun-04	Low flow	0	0	0	0	0	0	0	0
Puente la Victoria	24-Jun-04	Low flow	0	0	0	0	0	0	0	0
Anacaro	24-Jun-04	Low flow	0	1	0	0	12	0	0	0

Note: sp.: specie

Without. Id. sp: Specie without Identication

Annex F. Database used in this research for developing predictive macroinvertebrate models . (cont.)

Table F.1 Macroinvertebrate abundance database (cont.)

Sampling site	Date	Number of Organism collected								
		Díptera					Neuroptera	Ephemeróptera		
		Blepharoceride	Chironomidae	Syrphidae	Tipulidae	Simuliidae	Corydalidae	Leptohyphidae	Leptophlebiidae	Baetidae
Paso de la Balsa	19-Feb-97	0	0	0	0	0	6	0	14	13
Paso de la Bolsa	19-Feb-97	0	0	0	3	18	12	0	0	0
Juanchito	19-Feb-97	7	0	0	22	16	0	0	0	0
Paso de la Torre	19-Feb-97	0	0	22	0	0	0	0	0	0
Mediacanoa	19-Feb-97	0	0	0	0	22	0	0	0	0
Anacaro	25-Jul-01	3	0	0	0	0	0	0	0	0
Paso de la Balsa	22-Jun-04	1	0	0	0	0	0	4	5	0
Paso de la Bolsa	22-Jun-04	0	0	0	0	0	0	6	1	5
Puente Hormiguero	22-Jun-04	0	0	0	0	0	0	1	0	3
Antes Navarro	23-Jun-04	0	164	0	0	0	0	2	0	2
Juanchito	23-Jun-04	0	218	0	0	0	0	0	0	0
Paso de la Torre	23-Jun-04	0	372	1	0	0	0	0	0	0
Mediacanoa	23-Jun-04	0	65	0	0	0	0	0	0	0
Puente la Victoria	24-Jun-04	1	56	0	0	0	0	3	0	0
Anacaro	24-Jun-04	0	0	0	0	0	0	0	0	0

Note: sp.: specie
Without. Id. sp: Specie without Identication

Annex F. Database used in this research for developing predictive macroinvertebrate models. (cont.)

Table F.1 Macroinvertebrate abundance database (cont.)

Sampling site	Date	Number of Organism collected							
		Hemíptera			Odonata				
		Veliidae	Naucoridae	Gerridae	Libellulidae	Gomphidae (Phyllocycla sp.)	Gomphidae (Aphilla sp.)	Aesnidae	Gomphidae (Perigomphus sp.)
Paso de la Balsa	19-Feb-97	0	0	0	2	0	0	5	0
Paso de la Bolsa	19-Feb-97	0	0	0	11	0	0	0	0
Juanchito	19-Feb-97	0	0	0	5	0	0	0	0
Paso de la Torre	19-Feb-97	0	0	0	0	0	0	0	0
Mediacanoa	19-Feb-97	0	0	0	0	0	0	0	0
Anacaro	25-Jul-01	0	0	6	17	0	0	0	0
Paso de la Balsa	22-Jun-04	24	0	0	0	0	0	0	0
Paso de la Bolsa	22-Jun-04	16	1	0	1	0	0	0	0
Puente Hormiguero	22-Jun-04	0	1	0	0	0	0	0	0
Antes Navarro	23-Jun-04	0	0	0	0	0	0	0	0
Juanchito	23-Jun-04	10	0	0	0	0	0	0	0
Paso de la Torre	23-Jun-04	0	0	0	0	0	0	0	0
Mediacanoa	23-Jun-04	0	0	0	1	0	0	0	0
Puente la Victoria	24-Jun-04	0	1	0	0	0	1	0	1
Anacaro	24-Jun-04	6	0	0	0	1	0	0	1

Note: sp.: specie

Without. Id. sp: Specie without Identication

Annex F. Database used in this research for developing predictive macroinvertebrate models. (cont.)

Table F.1 Macroinvertebrate abundance database (cont.)

Sampling site	Date	Number of Organism collected							
		Trichóptera					Plecoptera	Mesogastrópoda	
		Leptoceridae (Atanatolica sp.)	Leptoceridae (Grumichella sp.)	Leptoceridae (Nectopsyche sp.)	Leptoceridae (Oecetis sp.)	Hydropsychidae	Perlidae	Thiaridae (Hemisinus sp.)	Thiaridae (Without. Id. sp)
Paso de la Balsa	19-Feb-97	0	0	0	0	10	0	0	0
Paso de la Bolsa	19-Feb-97	23	0	0	0	0	0	0	0
Juanchito	19-Feb-97	0	0	0	0	23	0	0	0
Paso de la Torre	19-Feb-97	0	0	0	0	0	0	0	0
Mediacanoa	19-Feb-97	0	0	0	0	0	0	0	0
Anacaro	25-Jul-01	0	0	2	0	4	13	0	0
Paso de la Balsa	22-Jun-04	4	1	0	0	2	0	2	0
Paso de la Bolsa	22-Jun-04	0	0	2	0	1	0	2	0
Puente Hormiguero	22-Jun-04	0	0	0	1	0	0	97	0
Antes Navarro	23-Jun-04	0	0	0	0	0	0	80	0
Juanchito	23-Jun-04	0	0	0	0	0	0	670	0
Paso de la Torre	23-Jun-04	0	0	0	0	0	0	45	0
Mediacanoa	23-Jun-04	0	0	0	0	0	0	67	0
Puente la Victoria	24-Jun-04	0	0	0	0	0	0	970	0
Anacaro	24-Jun-04	0	0	0	0	0	0	329	1020

Note: sp.: specie

Without. Id. sp: Specie without Identication

Annex F. Database used in this research for developing predictive macroinvertebrate models. (cont.)

Table F.1 Macroinvertebrate abundance database (cont.)

Sampling site	Date	Number of Organism collected								
		Haplotaxida				Tricladida	Basommatóphora		Glossiphoniiformes	Unionoidea
		Tubificidae	Without. Id. Sp 1.	Without. Id. Sp 2.	Without. Id. Sp 3.	Planariidae	Ancyliidae	Physidae	Glossiphoniidae	Mycetopodidae
Paso de la Balsa	19-Feb-97	0	0	0	0	0	0	0	0	0
Paso de la Bolsa	19-Feb-97	0	0	0	0	0	0	0	0	0
Juanchito	19-Feb-97	0	0	0	0	0	0	0	0	0
Paso de la Torre	19-Feb-97	53	0	0	0	22	0	66	0	0
Mediacanoa	19-Feb-97	58	0	0	0	0	0	0	0	0
Anacaro	25-Jul-01	0	0	0	0	0	0	0	0	0
Paso de la Balsa	22-Jun-04	0	0	0	0	0	0	0	0	0
Paso de la Bolsa	22-Jun-04	0	0	0	0	0	0	0	0	0
Puente Hormiguero	22-Jun-04	500	0	0	0	2	0	0	0	0
Antes Navarro	23-Jun-04	752	0	0	0	0	41	0	0	0
Juanchito	23-Jun-04	7500	1	8	1	0	0	0	0	0
Paso de la Torre	23-Jun-04	32500	1	4	0	0	0	0	0	0
Mediacanoa	23-Jun-04	6000	0	1	0	0	0	1	2	0
Puente la Victoria	24-Jun-04	127	0	0	0	0	0	0	13	3
Anacaro	24-Jun-04	110	85	0	0	8	45	0	2	0

Note: sp.: specie

Without. Id. sp: Specie without Identication

Annex F. Database used in this research for developing predictive macroinvertebrate models. (cont.)

Table F.2 Macroinvertebrate occurrence database. Record of presence (1) or absence (0)

Sampling site	Date	Condition	Presence (1) and Absence (0) of organism							
			Coleóptera							
			Psephenidae	Elmidae (Macrelmis sp.)	Elmidae (Cylloepus sp.)	Elmidae (Heterelmis sp.)	Elmidae (Without. Id. sp)	Staphylinidae	Ptylodactylidae	Hydrophilidae
Paso de la Balsa	19-Feb-97	High flow	0	1	0	0	0	0	0	0
Paso de la Bolsa	19-Feb-97	High flow	0	0	0	0	0	0	0	0
Juanchito	19-Feb-97	High flow	0	0	0	0	0	0	0	0
Paso de la Torre	19-Feb-97	High flow	0	0	0	0	0	0	0	0
Mediacanoa	19-Feb-97	High flow	0	0	1	0	0	0	1	0
Anacaro	25-Jul-01	Low flow	0	0	0	0	0	0	1	0
Paso de la Balsa	22-Jun-04	Low flow	1	1	0	0	0	1	0	0
Paso de la Bolsa	22-Jun-04	Low flow	0	0	0	1	0	0	0	0
Puente Hormiguero	22-Jun-04	Low flow	0	1	0	0	0	1	0	0
Antes Navarro	23-Jun-04	Low flow	0	0	0	1	0	0	0	0
Juanchito	23-Jun-04	Low flow	0	0	0	0	0	0	0	0
Paso de la Torre	23-Jun-04	Low flow	0	0	0	0	0	0	0	1
Mediacanoa	23-Jun-04	Low flow	0	0	0	0	0	0	0	0
Puente la Victoria	24-Jun-04	Low flow	0	0	0	0	0	0	0	0
Anacaro	24-Jun-04	Low flow	0	1	0	0	1	0	0	0

Note: sp.: specie
Without. Id. sp: Specie without Identication

Annex F. Database used in this research for developing predictive macroinvertebrate models. (cont.)

Table F.2 Macroinvertebrate occurrence database. Record of presence (1) or absence (0). (cont.)

Sampling site	Date	Presence (1) and Absence (0) of organism								
		Díptera					Neuroptera	Ephemeróptera		
		Blepharoceride	**Chironomidae**	**Syrphidae**	**Tipulidae**	**Simuliidae**	**Corydalidae**	**Leptohyphidae**	**Leptophlebiidae**	**Baetidae**
Paso de la Balsa	19-Feb-97	0	0	0	0	0	1	0	1	1
Paso de la Bolsa	19-Feb-97	0	0	0	1	1	1	0	0	0
Juanchito	19-Feb-97	1	0	0	1	1	0	0	0	0
Paso de la Torre	19-Feb-97	0	0	1	0	0	0	0	0	0
Mediacanoa	19-Feb-97	0	0	0	0	1	0	0	0	0
Anacaro	25-Jul-01	1	0	0	0	0	0	0	0	0
Paso de la Balsa	22-Jun-04	1	0	0	0	0	0	1	1	0
Paso de la Bolsa	22-Jun-04	0	0	0	0	0	0	1	1	1
Puente Hormiguero	22-Jun-04	0	0	0	0	0	0	1	0	1
Antes Navarro	23-Jun-04	0	1	0	0	0	0	1	0	1
Juanchito	23-Jun-04	0	1	0	0	0	0	0	0	0
Paso de la Torre	23-Jun-04	0	1	1	0	0	0	0	0	0
Mediacanoa	23-Jun-04	0	1	0	0	0	0	0	0	0
Puente la Victoria	24-Jun-04	1	1	0	0	0	0	1	0	0
Anacaro	24-Jun-04	0	0	0	0	0	0	0	0	0

Note: sp.: specie
Without. Id. sp: Specie without Identication

Annex F. Database used in this research for developing predictive macroinvertebrate models. (cont.)

Table F.2 Macroinvertebrate occurrence database. Record of presence (1) or absence (0). (cont.)

Sampling site	Date	Presence (1) and Absence (0) of organism							
		Hemíptera			Odonata				
		Veliidae	Naucoridae	Gerridae	Libellulidae	Gomphidae (Phyllocycla sp.)	Gomphidae (Aphilla sp.)	Aesnidae	Gomphidae (Perigomphus sp.)
Paso de la Balsa	19-Feb-97	0	0	0	1	0	0	1	0
Paso de la Bolsa	19-Feb-97	0	0	0	1	0	0	0	0
Juanchito	19-Feb-97	0	0	0	1	0	0	0	0
Paso de la Torre	19-Feb-97	0	0	0	0	0	0	0	0
Mediacanoa	19-Feb-97	0	0	0	0	0	0	0	0
Anacaro	25-Jul-01	0	0	1	1	0	0	0	0
Paso de la Balsa	22-Jun-04	1	0	0	0	0	0	0	0
Paso de la Bolsa	22-Jun-04	1	1	0	1	0	0	0	0
Puente Hormiguero	22-Jun-04	0	1	0	0	0	0	0	0
Antes Navarro	23-Jun-04	0	0	0	0	0	0	0	0
Juanchito	23-Jun-04	1	0	0	0	0	0	0	0
Paso de la Torre	23-Jun-04	0	0	0	0	0	0	0	0
Mediacanoa	23-Jun-04	0	0	0	1	0	0	0	0
Puente la Victoria	24-Jun-04	0	1	0	0	0	1	0	1
Anacaro	24-Jun-04	1	0	0	0	1	0	0	1

Note: sp.: specie
Without. Id. sp: Specie without Identication

Annex F. Database used in this research for developing predictive macroinvertebrate models. (cont.)

Table F.2 Macroinvertebrate occurrence database. Record of presence (1) or absence (0). (cont.)

Sampling site	Date	Presence (1) and Absence (0) of organism							
		Trichóptera					Plecoptera	Mesogastrópoda	
		Leptoceridae (Atanatolica sp.)	Leptoceridae (Grumichella sp.)	Leptoceridae (Nectopsyche sp.)	Leptoceridae (Oecetis sp.)	Hydropsychidae	Perlidae	Thiaridae (Hemisinus sp.)	Thiaridae (Without. Id. sp)
Paso de la Balsa	19-Feb-97	0	0	0	0	1	0	0	0
Paso de la Bolsa	19-Feb-97	1	0	0	0	0	0	0	0
Juanchito	19-Feb-97	0	0	0	0	1	0	0	0
Paso de la Torre	19-Feb-97	0	0	0	0	0	0	0	0
Mediacanoa	19-Feb-97	0	0	0	0	0	0	0	0
Anacaro	25-Jul-01	0	0	1	0	1	1	0	0
Paso de la Balsa	22-Jun-04	1	1	0	0	1	0	1	0
Paso de la Bolsa	22-Jun-04	0	0	1	0	1	0	1	0
Puente Hormiguero	22-Jun-04	0	0	0	1	0	0	1	0
Antes Navarro	23-Jun-04	0	0	0	0	0	0	1	0
Juanchito	23-Jun-04	0	0	0	0	0	0	1	0
Paso de la Torre	23-Jun-04	0	0	0	0	0	0	1	0
Mediacanoa	23-Jun-04	0	0	0	0	0	0	1	0
Puente la Victoria	24-Jun-04	0	0	0	0	0	0	1	0
Anacaro	24-Jun-04	0	0	0	0	0	0	1	1

Note: sp.: specie
Without. Id. sp: Specie without Identication

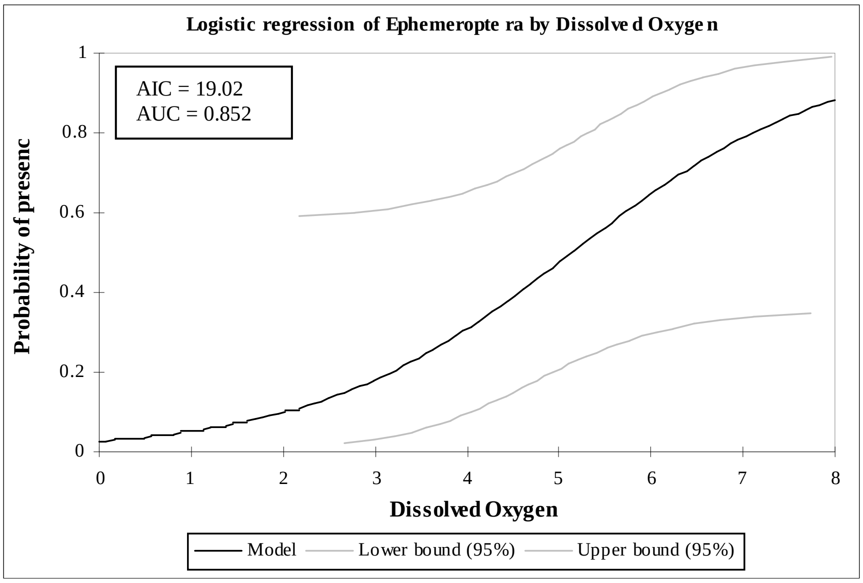
Annex F. Database used in this research for developing predictive macroinvertebrate models. (cont.)

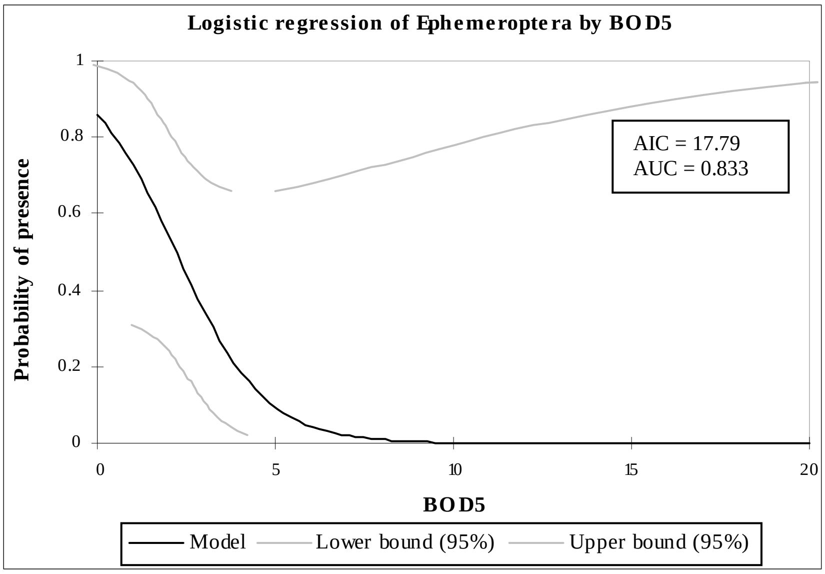
Table F.2 Macroinvertebrate occurrence database. Record of presence (1) or absence (0). (cont.)

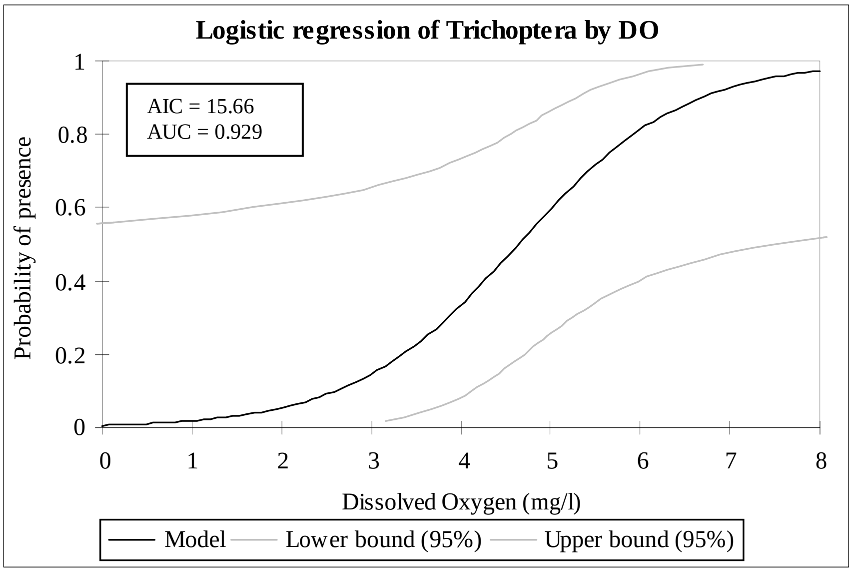
Sampling site	Date	Presence (1) and Absence (0) of organism								
		Haplotaxida				**Tricladida**	**Basommatóphora**		**Glossiphoniiformes**	**Unionoidea**
		Tubificidae	**Without. Id. Sp 1.**	**Without. Id. Sp 2.**	**Without. Id. Sp 3.**	**Planariidae**	**Ancyliidae**	**Physidae**	**Glossiphoniidae**	**Mycetopodidae**
Paso de la Balsa	19-Feb-97	0	0	0	0	0	0	0	0	0
Paso de la Bolsa	19-Feb-97	0	0	0	0	0	0	0	0	0
Juanchito	19-Feb-97	0	0	0	0	0	0	0	0	0
Paso de la Torre	19-Feb-97	1	0	0	0	1	0	1	0	0
Mediacanoa	19-Feb-97	1	0	0	0	0	0	0	0	0
Anacaro	25-Jul-01	0	0	0	0	0	0	0	0	0
Paso de la Balsa	22-Jun-04	0	0	0	0	0	0	0	0	0
Paso de la Bolsa	22-Jun-04	0	0	0	0	0	0	0	0	0
Puente Hormiguero	22-Jun-04	1	0	0	0	1	0	0	0	0
Antes Navarro	23-Jun-04	1	0	0	0	0	1	0	0	0
Juanchito	23-Jun-04	1	1	1	1	0	0	0	0	0
Paso de la Torre	23-Jun-04	1	1	1	0	0	0	0	0	0
Mediacanoa	23-Jun-04	1	0	1	0	0	0	1	1	0
Puente la Victoria	24-Jun-04	1	0	0	0	0	0	0	1	1
Anacaro	24-Jun-04	1	1	0	0	1	1	0	1	0

Note: sp.: specie
Without. Id. sp: Specie without Identication

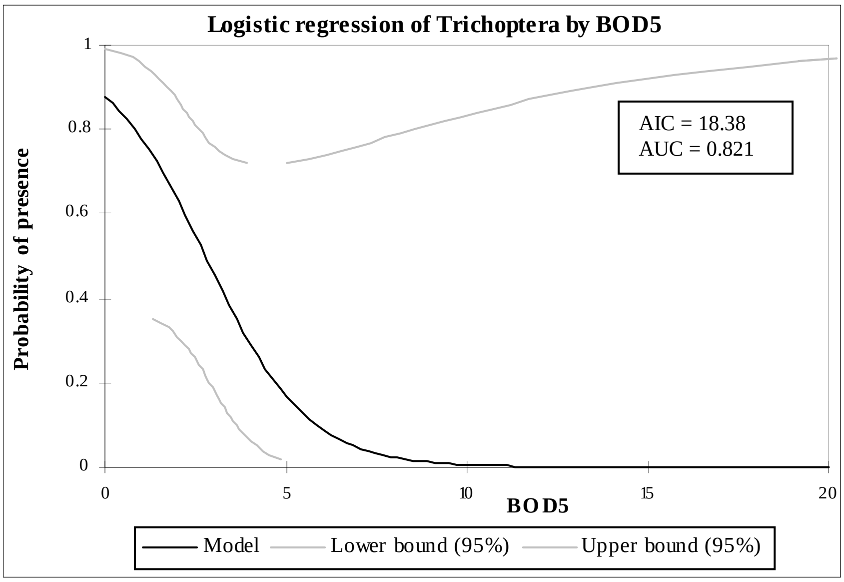
Annex G. Outputs of the software XLSTAT for Simple logistic regression models

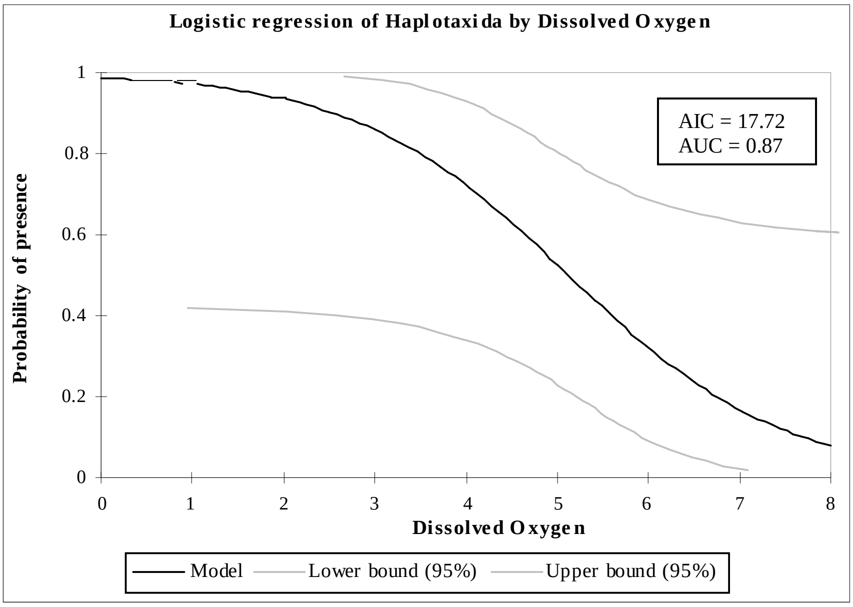


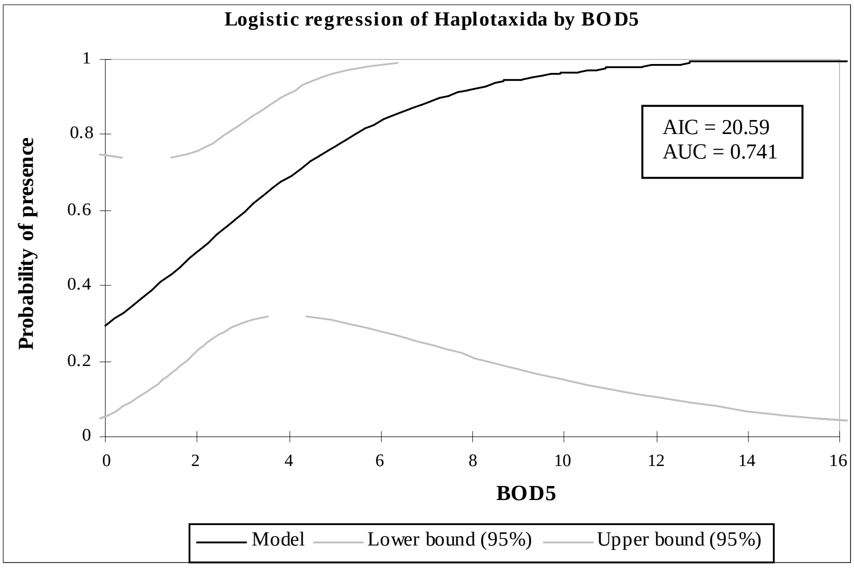




Annex G. Outputs of the software XLSTAT for Simple logistic regression models. (cont.)





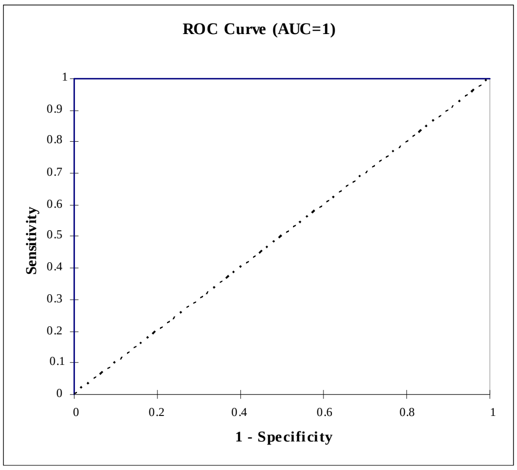


Annex H Output of the software XLSTAT for the best MLRMs selected.

H.1 Ephemeroptera: MLRM with DO, flow and depth

Goodness of fit statistics (Variable Ephemeroptera):

Statistic	Independent	Full
Observations	15	15
Sum of weights	15.000	15.000
DF	14	11
-2 Log(Likelihood)	20.190	0.000
R²(McFadden)	0.000	1.000
R²(Cox and Snell)	0.000	0.740
R²(Nagelkerke)	0.000	1.000
AIC	22.190	8.000
SBC	22.898	10.832
Iterations	0	20



Test of the null hypothesis H0: Y=0.400 (Variable Ephemeroptera)

Statistic	DF	Chi-square	Pr > Chi²
-2 Log(Likelihood)	3	20.190	0.0002
Score	3	8.914	0.030
Wald	3	0.000	1.000

Type III analysis (Variable Ephemeroptera)

Source	DF	Chi-square (Wald)	Pr > Wald	Chi-square (LR)	Pr > LR
Flow	1	-10841684938113600.000	1.000	432.524	< 0.0001
Depth	1	-16100282810853700.000	1.000	432.524	< 0.0001
Dissolved Oxygen	1	-19672099765427100.000	1.000	432.524	< 0.0001

Equation of the model (Variable Ephemeroptera):

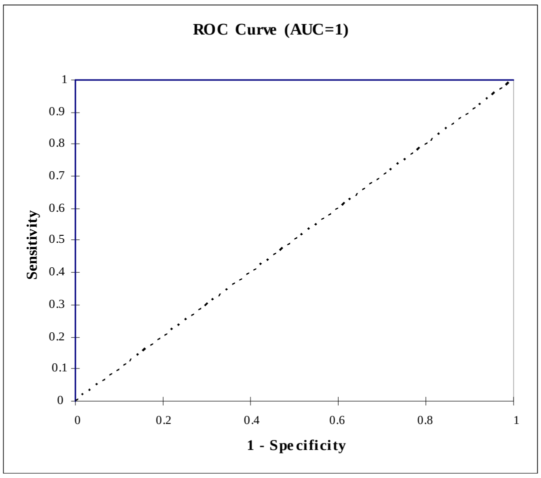
Ephemeroptera = 1 / (1 + exp(-(207.899423834621-0.820551295778075*Flow-83.9681211412539*Depth+64.9079354994629*Dissolved Oxygen)))

Annex H Output of the software XLSTAT for the best MLRMs selected. (cont.)

H.2 Trichoptera: MLRM with BOD_5, velocity, flow and depth

Goodness of fit statistics (Variable Trichoptera):

Statistic	Independent	Full
Observations	15	15
Sum of weights	15.000	15.000
DF	14	10
-2 Log(Likelihood)	20.728	0.000
R²(McFadden)	0.000	1.000
R²(Cox and Snell)	0.000	0.749
R²(Nagelkerke)	0.000	1.000
AIC	22.728	10.000
SBC	23.436	13.540
Iterations	0	21



Test of the null hypothesis H0: Y=0.467 (Variable Trichoptera)

Statistic	DF	Chi-square	Pr > Chi²
-2 Log(Likelihood)	4	20.728	0.0004
Score	4	7.132	0.129
Wald	4	0.000	1.000

Type III analysis (Variable Trichoptera)

Source	DF	Chi-square (Wald)	Pr > Wald	Chi-square (LR)	Pr > LR
Flow	1	-17506968176065100.000	1.000	360.366	< 0.0001
Depth	1	-4448745486994550.000	1.000	178.295	< 0.0001
Velocity	1	-452737647574515.000	1.000	71.378	< 0.0001
BOD5	1	-116337839243146000.000	1.000	467.511	< 0.0001

Equation of the model (Variable Trichoptera):

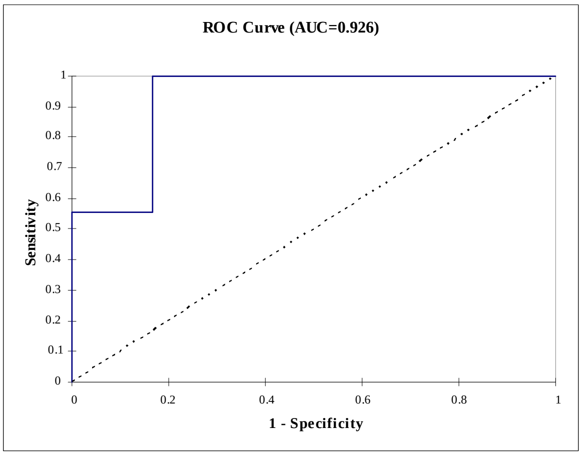
Trichoptera = 1 / (1 + exp(-(121.646115255732+1.04270826400976*Flow-44.1383875332197*Depth+103.421063163048*Velocity-79.3010221223499*BOD5)))

Annex H Output of the software XLSTAT for the best MLRMs selected. (cont.)

H.3 Haplotaxida: MLRM with BOD_5 and velocity

Goodness of fit statistics (Variable Haplotaxida):

Statistic	Independent	Full
Observations	15	15
Sum of weights	15.000	15.000
DF	14	12
-2 Log(Likelihood)	20.190	11.614
R²(McFadden)	0.000	0.425
R²(Cox and Snell)	0.000	0.435
R²(Nagelkerke)	0.000	0.589
AIC	22.190	17.614
SBC	22.898	19.738
Iterations	0	7



Test of the null hypothesis H0: Y=0.600 (Variable Haplotaxida)

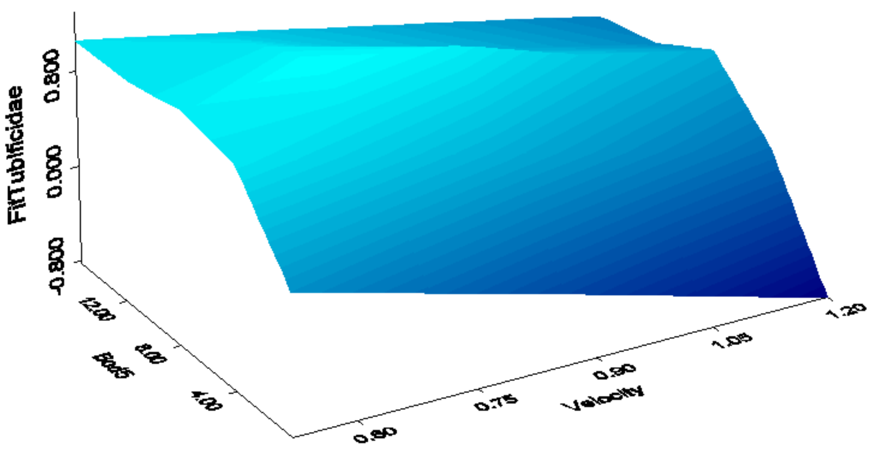
Statistic	DF	Chi-square	Pr > Chi²
-2 Log(Likelihood)	2	8.577	0.014
Score	2	4.365	0.113
Wald	2	3.366	0.186

Type III analysis (Variable Haplotaxida)

Source	DF	Chi-square (Wald)	Pr > Wald	Chi-square (LR)	Pr > LR
Velocity	1	2.614	0.106	4.980	0.026
BOD5	1	3.021	0.082	6.482	0.011

Equation of the model (Variable Haplotaxida):

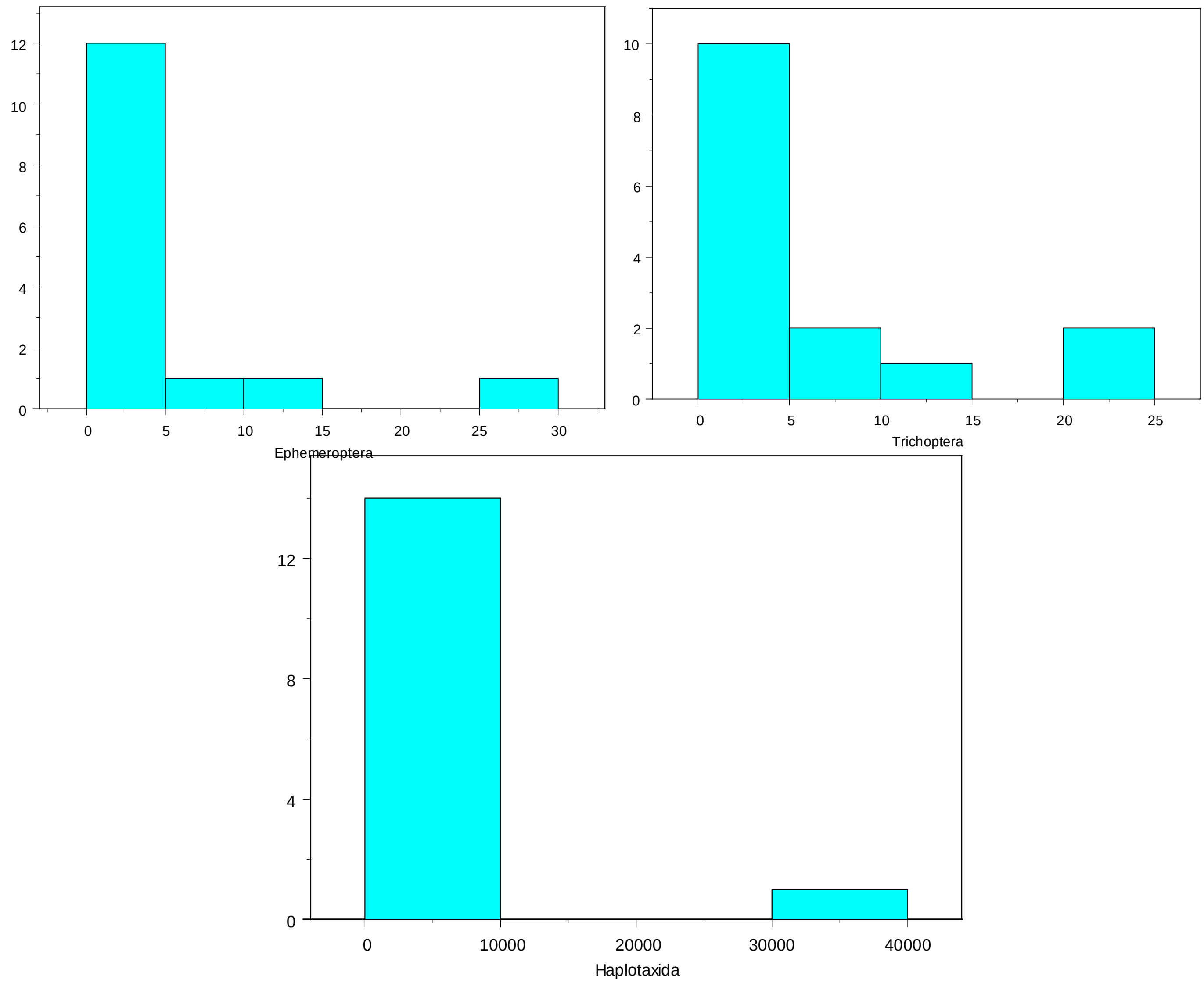
Tubificidae = 1 / (1 + exp(-(4.26050001421182-
9.31956001664771*Velocity+1.06018574795165*BOD5)))



Annex I. Statistical analysis of the abundance data used for estimating QPRM

	Ephemeroptera	Trichoptera	Haplotaxida
Min:	0.0000000	0.0000000	0.000000e+000
1st Qu.:	0.0000000	0.0000000	0.000000e+000
Mean:	3.9333333	4.8666667	3.180067e+003
Median:	0.0000000	0.0000000	5.800000e+001
3rd Qu.:	4.0000000	6.5000000	6.260000e+002
Max:	27.0000000	23.0000000	3.250500e+004
Total N:	15.0000000	15.0000000	1.500000e+001
NA's :	0.0000000	0.0000000	0.000000e+000
Variance:	54.4952381	64.1238095	7.129790e+007
Std Dev.:	7.3820890	8.0077344	8.443808e+003
Sum:	59.0000000	73.0000000	4.770100e+004
SE Mean:	1.9060472	2.0675881	2.180182e+003
LCL Mean:	-0.1547313	0.4321312	-1.495959e+003
UCL Mean:	8.0213980	9.3012021	7.856092e+003
Skewness:	2.5362283	1.7792282	3.419644e+000
Kurtosis:	6.9672240	2.1552597	1.224908e+001

Histograms

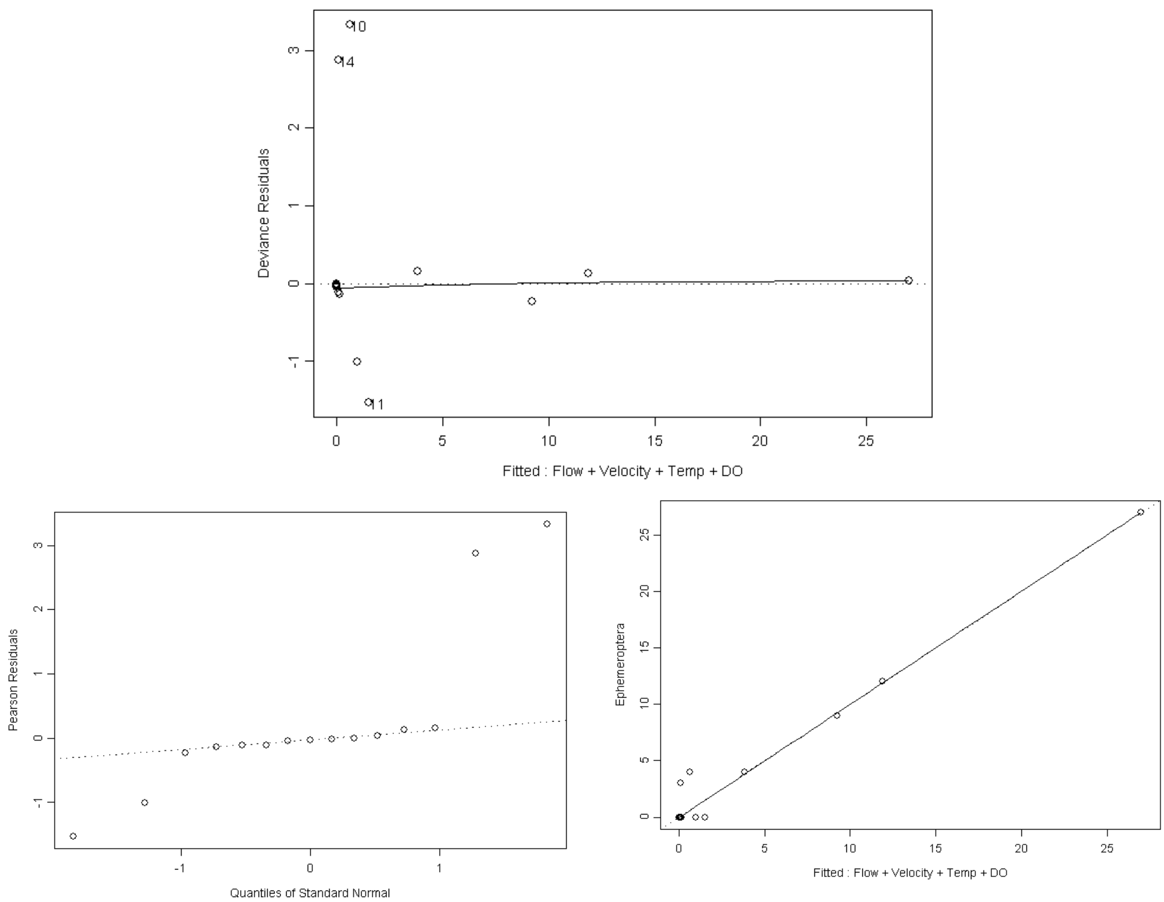


Annex J. Output of the software XLSTAT for the best QPRMs selected. (cont.)

J.1 Ephemeroptera: QPRM with flow, velocity, temperature and DO

*** Generalized Linear Model ***

```
Call: glm(formula = Ephemeroptera ~ Flow + Velocity + Temp + DO, family = quasi(
       link = log, variance = "mu"), data = DBmonitoringComplete,
       na.action = na.exclude, control = list(epsilon = 0.0001, maxit = 50,
       trace = F))
Deviance Residuals:
       Min         1Q      Median          3Q       Max
 -1.523538 -0.1272909 -0.02449327 0.07791923 3.329471

Coefficients:
                   Value Std. Error    t value
(Intercept)  12.63569149 8.66870800  1.4576211
       Flow  -0.02234925 0.01528909 -1.4617776
   Velocity   3.09557983 2.48170514  1.2473600
       Temp  -0.57976769 0.25080997 -2.3115815
         DO   0.49883030 0.50696273  0.9839585

(Dispersion Parameter for Quasi-likelihood family taken to be 2.285481 )

    Null Deviance: 762.9333 on 14 degrees of freedom

Residual Deviance: 22.85509 on 10 degrees of freedom

Number of Fisher Scoring Iterations: 7
```

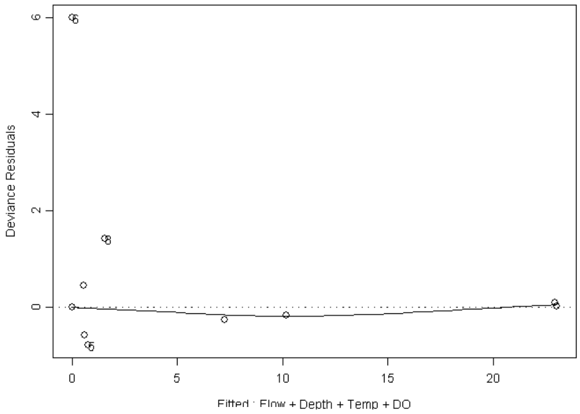


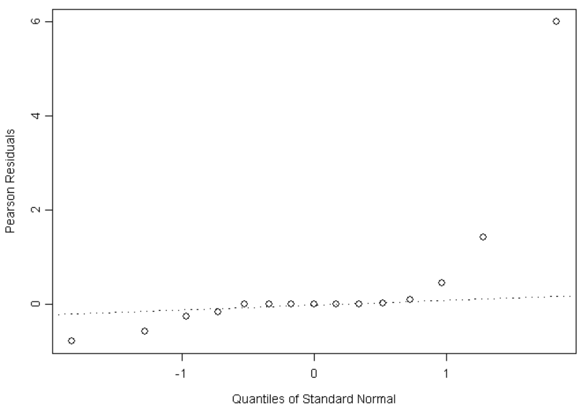
Annex J. Output of the software XLSTAT for the best QPRMs selected. (cont.)

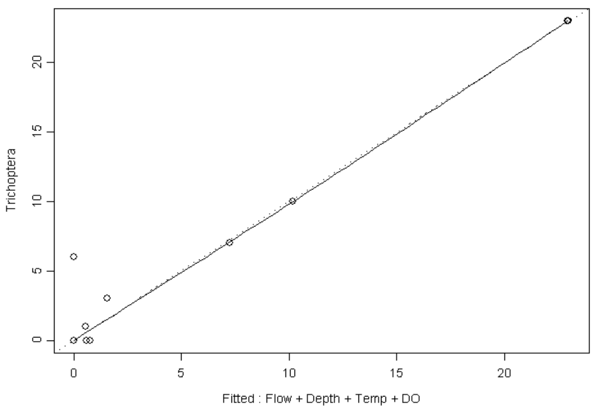
J.2 Trichoptera: QPRM with flow, depth, temperature and DO

*** Generalized Linear Model ***

```
Call: glm(formula = Trichoptera ~ Flow + Depth + Temp + DO, family = quasi(link =
       log, variance = "mu"), data = DBmonitoringComplete, na.action =
       na.exclude, control = list(epsilon = 0.0001, maxit = 50, trace = F))
Deviance Residuals:
        Min          1Q         Median        3Q       Max
 -0.7793026 -0.09187625 -1.332815e-007 0.0462149 5.999969

Coefficients:
                  Value Std. Error   t value
(Intercept) -58.6904482 21.2132097 -2.766693
       Flow   0.0486287  0.0149081  3.261898
      Depth  -0.8992582  0.2555749 -3.518570
       Temp   1.1569406  0.3831915  3.019223
         DO   4.6339058  1.7021407  2.722399

(Dispersion Parameter for Quasi-likelihood family taken to be 3.930323)

    Null Deviance: 897.7333 on 14 degrees of freedom

Residual Deviance: 39.30335 on 10 degrees of freedom

Number of Fisher Scoring Iterations: 8
```



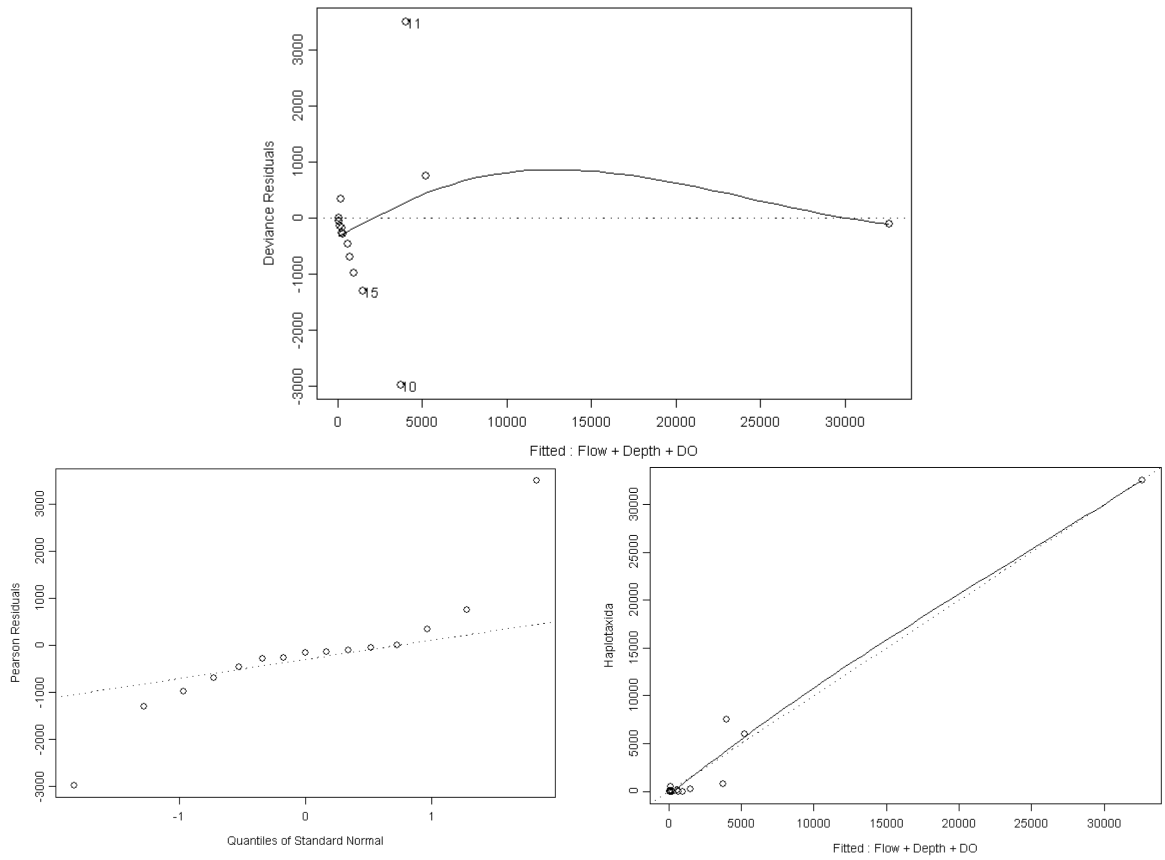




Annex J. Output of the software XLSTAT for the best QPRMs selected. (cont.)

J.3 Haplotaxida: QPRM with flow, depth and DO

```
                         *** Generalized Linear Model ***

Call: glm(formula = Haplotaxida ~ Flow + Depth + DO, family = quasi(link = log,
       variance = "mu"), data = DBmonitoringComplete, na.action =
       na.exclude, control = list(epsilon = 0.0001, maxit = 50, trace = F))
Deviance Residuals:
       Min       1Q    Median        3Q      Max
 -2970.286 -571.482 -165.1311 -22.33457 3506.822

Coefficients:
                  Value Std. Error   t value
(Intercept)  7.91158512 3.08436112  2.565064
       Flow -0.01710292 0.01222058 -1.399518
      Depth  0.98501584 0.36924317  2.667662
         DO -0.47932747 0.06580388 -7.284183

(Dispersion Parameter for Quasi-likelihood family taken to be 2300834 )

    Null Deviance: 164820 on 14 degrees of freedom

Residual Deviance: 12184.47 on 11 degrees of freedom

Number of Fisher Scoring Iterations: 5
```



Annex K. Output of the software XLSTAT for the QPRMs for the BMWP index

```
                        *** Generalized Linear Model ***

Call: glm(formula = BMWP ~ Flow + Depth + Velocity + Temp + DO, family = quasi(
       link = log, variance = "mu"), data = DBmonitoringBMWP, na.action =
       na.exclude, control = list(epsilon = 0.0001, maxit = 50, trace = F))
Deviance Residuals:
       Min        1Q     Median       3Q     Max
 -3.934386 -1.461115 -0.2700082 1.324239 3.07984

Coefficients:
                   Value  Std. Error    t value
(Intercept) -3.639163910 3.774711940 -0.9640905
       Flow  0.002277795 0.002064421  1.1033578
      Depth -0.289517955 0.135169924 -2.1418815
   Velocity  2.088739247 1.453266519  1.4372720
       Temp  0.228546637 0.124859378  1.8304323
         DO  0.228757311 0.092758916  2.4661490

(Dispersion Parameter for Quasi-likelihood family taken to be 6.06427 )

    Null Deviance: 153.9031 on 14 degrees of freedom

Residual Deviance: 57.50741 on 9 degrees of freedom

Number of Fisher Scoring Iterations: 4
```