Development and application of an integrated ecological modelling framework to analyze the impact of wastewater discharges on the ecological water quality of rivers

Javier E. Holguin-Gonzalez a, Gert Everaert a, Pieter Boets a,*, Alberto Galvis b, Peter L.M. Goethals a

a Ghent University, Laboratory of Environmental Toxicology and Aquatic Ecology, J. Plateaustraat 22, B-9000 Ghent, Belgium
b Universidad del Valle, Instituto Cinara, Calle 13 # 100-00 Ed. 341, Cali, Colombia

Abstract

Modelling is an effective tool to investigate the ecological state of water resources. In developing countries, the impact of sanitation infrastructures (e.g. wastewater treatment plants) is typically assessed considering the achievement of legal physicochemical quality standards, but ignoring the ecological water quality (EWQ) of the receiving river. In this paper, we developed a generic integrated ecological modelling framework quantifying the impact of wastewater discharges on the EWQ of the Cauca river (Colombia). The framework is flexible enough to be used in conjunction with different approaches/models and integrates a hydraulic and physicochemical water quality model with aquatic ecological models. Two types of ecological models were developed, habitat suitability models for selected macroinvertebrate groups and ecological assessment models based on a macroinvertebrate biotic index. Four pollution control scenarios were tested. It was found that the foreseen investments in sanitation infrastructure will lead to modest improvements of the EWQ, with an increase lower than six units of the ecological index BMWP-Colombia. Advanced investments, such as the collection and treatment of all wastewater produced by the cities of Cali, Yumbo and Palmira and upgrading of the treatment systems should be considered to achieve a good EWQ. The results show that the integration of ecological models in hydraulic and physicochemical water quality models (e.g. MIKE 11) has an added value for decision support in river management and water policy. The integration of models is a key aspect for the success in environmental decision making. The main limitation of this approach is the availability of physicochemical, hydraulic and biological data that are collected simultaneously. Therefore, a change in the river monitoring strategy towards collection of data which include simultaneous measurements of these variables is required.

1. Introduction

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

During the last decade, the integration of hydro-morphological, physicochemical and ecological models for decision support in river management started gaining interest (Mouton et al., 2009; Vaughan et al., 2009; Hughes and Louw, 2010; Boets et al., 2013). From an ecological point of view, benthic macroinvertebrates have been chosen as ecological indicators because they are expected to respond to both physicochemical and hydro-morphological pressures, and can act as a link between primary producers and higher organisms (De Pauw and Hawkes, 1993; De Pauw et al., 2006). Recently, researchers emphasized on the integration of hydraulic/hydrodynamic models with habitat suitability index (HSI) curves for macroinvertebrates (e.g. Bockelmann et al., 2004; Tomsic et al., 2007). This HSI approach considers hydro-morphological pressures (e.g. changes in water depth, water velocity, type of substrate), but omits the impact of physicochemical pressures (i.e. physicochemical pollution). More recently, Mouton et al. (2009) considered the impact of these two types of pressures on the ecological river quality, with an application of the Water Framework Directive Explorer (WFD-Explorer) toolbox. The WFD-Explorer includes a one-dimensional hydraulic model linked to a mass balance module that allowed them to predict the ecological water quality (EWQ) based on ecological expert knowledge rules. However, this toolbox simplifies water quality processes as a retention factor. Moreover, it operates at the coarse river basin scale level; whereas the impact of physical habitat changes on river biology occurs at smaller scale levels, such as mesoscale or microscale level (Mouton et al., 2009). Additionally, its knowledge rules were developed based on empirical data of Dutch and Flemish lowland streams, therefore, the transferability of these rules to other ecoregions in the world is limited (Randin et al., 2006; Fitzpatrick et al., 2007).

Considering the limitations of the HSI and WFD-Explorer approaches, there is a need for an integrated approach that allows us to assess simultaneously the impact of hydro-morphological pressures and physicochemical pollution on the ecological river water quality. This approach should include a detailed physical habitat and water quality model linked to ecological models based on specific characteristics of the studied river. In this research, a generic integrated ecological modelling framework for decision support in river management was proposed (Fig. 1), tested and validated on a case study of a lowland river basin in Colombia (Cauca river). The framework integrates a river water quantity and quality model with two types of ecological models, habitat suitability and ecological assessment models. This integrative framework was used to assess the ecological benefit of investments in sanitation infrastructure in the Cauca river by considering four pollution control scenarios and coupling the water quality model (MIKE 11 model; DHI, 1999) with EWQ models.

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

2. Materials and methods

2.1. Study area

The Cauca river is the second most important river in Colombia and the main hydrologic resource of southwest Colombia. The Cauca river’s valley is especially important for the country’s development and economy (CVC and Univalle, 2007). A significant part of the south-western manufacturing industry, the paper and sugar cane industry as well as part of the coffee producing zone are located along the river. The rapid urbanization and major economic development in the Cauca river’s valley, has led to dramatic degradation of the environment. There is an increasing deterioration of the water quality of this river due to wastewater discharges from domestic and industrial activities. This study focuses on the river stretch from the station Paso de La Balsa (abscissa 27.4 km and elevation of 965 metres above sea level m.a.s.l) to the station Anacaro (abscissa 416.5 km and 805 m.a.s.l) (Fig. 2) with a total length of 389.1 km. Multiple water quality problems can be found in this zone, especially in the dry season, downstream from the cities of Cali, Yumbo and Palmira (main industrial cities in the region). Under low flow conditions the Biological Oxygen Demand (BODs) and Faecal Coliforms can rise up to 7.5 mg/L and 2.4 * 10^6 MPN/100 mL, respectively, whereas the Dissolved Oxygen (DO) concentration can drop near to 0 mg/L. The city of Cali, with more than two million inhabitants, is the main source of pollution as 60% of all wastewater does not receive any type of treatment and is directly discharged into the Cauca river (CVC and Univalle, 2007).

2.2. Data collection and dataset pre-processing

The dataset used in this research corresponds to the information collected in a 10 year period (1996–2005) by the CVC and the CRMP Project in the Cauca river (CVC and Univalle, 2007). Two types of datasets were used, the first one for the implementation of the MIKE 11 model and the second one for building the ecological models. Two monitoring campaigns with calibration and verification purposes for the MIKE 11 model were carried out during the years 2003 and 2005 considering hourly measurements during low and high flow conditions. For the ecological models, a dataset was developed which included simultaneous measurements (based on sampling location and time) of physicochemical data, hydraulic data and biological information. The biological information

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

MIKE 11 is a water quality model that predicts physicochemical variables under different water management scenarios. In order to enable the coupling between the ecological models and the MIKE 11 outcomes, only the six variables modelled by the MIKE 11 model (i.e. temperature, BOD₅, DO, flow, water depth and water velocity) and the biological information were retained. The final dataset for the ecological models consisted of these six variables (called predictors) and two response variables (presence/absence of macroinvertebrates and BMWP-Colombia values).

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

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

2.3. Water quality assessment

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

2.4. Water quality modelling techniques used in the catchment of the Cauca river

When dealing with model integration there are two general approaches that can be implemented. The integrative approach, in which new models are built for each application and the combinatorial approach which uses the existing models available (Goethals et al., 2007). However, there is an in-between approach combining existing and new models. In this research, the third approach was followed. The integrated ecological modelling framework proposed has three basic components (Fig. 1): (1) a river water quantity model, (2) a river water quality model and, (3) river habitat suitability and ecological assessment models. For the first and second components, the hydrodynamic and physicochemical water quality model MIKE 11 was used. For the third component, logistic regression (presence/absence predictions) and negative binomial regression (BMWP-Colombia index predictions) were implemented. Once the integration of models is performed, they can be used for model simulations. The ecological models developed were applied on the resulting hydraulic and physicochemical data of the water quality scenarios generated by simulations with the MIKE 11 model. An overview of the modelling techniques and different modelling processes implemented is presented in Table 1.

2.4.1. River water quantity and quality model

The hydrodynamic and physicochemical water quality model MIKE 11 (DHI, 1999) used in this research is a mathematical simulation model which was calibrated and verified for dynamic flow conditions in the framework of the CRMP Project (CVC and Univalle, 2007). The water quality modelling by the MIKE 11 model included temperature, BOD₅ and DO as state variables. Details about the configuration of the MIKE 11 model in the Cauca river are described by Holguin (2007). The results of the calibration and validation of the MIKE 11 model can be analyzed in two ways. The first analysis considers hourly variation of the physicochemical variables in each station during the monitoring days and the second analysis consists of an instantaneous profile of the values of the variables in all the stations simultaneously. In this study, we focused on the first analysis, which gives a better idea of the modelling output under dynamic conditions.

A sensitivity analysis, based on the parameter perturbation method (Chapra, 1997), was performed to select the most sensitive calibration parameters. The re-aeration formula proposed by O'Connor and Dobbins (1956) gave the best correlations with the experimental re-aeration rates obtained during the CRMP Project. For the calibration and uncertainty estimation the concepts of the Generalised Likelihood Uncertainty Estimation methodology (GLUE; Beven and Binley, 1992) were applied using the Monte Carlo analysis. The GLUE technique allows generating confidence bands for the model results. For the Monte Carlo analysis, thousands of combinations of the most sensitive calibration parameters, considering values from uniform distributions, were evaluated with simulations considering the data of the

| Table 1 Overview of the implemented modelling techniques, the different components of the model and the model building, validation, fitting and uncertainty (MSE: mean squared error, CCI: correctly classified instances, K: Cohen’s kappa coefficient, AUC: area under the receiver-operating-characteristic curve, r: Pearson correlation coefficient, r²: determination coefficient). |
| --- | --- | --- | --- | --- |
| **Model component** | **Model building** | **Model validation** | **Model fitting** | **Model uncertainty** |
| Water quality and quantity model (MIKE 11) | Monte Carlo simulations | Independent dataset | MSE | GLUE |
| Habitat suitability model (logistic regression) | Multi-model | Post hoc evaluation of the model adequacy and predictive performance | CCI, K | AUC |
| Ecological assessment model (negative binomial regression) | | | r, r² | |
monitoring campaign of 2005. The calibration ranges of these rates were selected considering the experimental values and ranges reported in the literature (Bowie et al., 1985; Chapra, 1997). The goodness of fit considered during the calibration was the Mean Squared Error (MSE). The MSE was calculated for each model run performed during the calibration and for each modelled variable. The model with the lowest MSE for the two variables (BOD; and DO) simultaneously was selected, leading to the best combination of values of the most sensitive calibration parameters. For the validation process the model was run using the data from 2003 without changing the calibrated parameters.

2.4.2. River habitat suitability and ecological assessment models

The approach followed for the ecological modelling in this research was to use multivariate statistical methods based on Generalized Linear Models (GLM). Parametric methods such as GLM are generally more efficient in small datasets than non-parametric methods such as Generalized Additive Models (GAM) or classification trees (Vayssières et al., 2000). GLM provide users with a conventional mathematical function and are better suited for analyzing ecological relationships, which can be poorly represented by classical Gaussian distributions (Zuur et al., 2007). Considering the experimental values and value ranges reported in the literature (Bowie et al., 1985; Chapra, 1997). The goodness of fit considered during the calibrations (Zuur et al., 2007).

The next step in the model building process is to identify the key explanatory variables for the LRM and the NBRM. Thereby, a multi-model inference technique based on information-theoretic (I-T) approach (Burnham and Anderson, 2002), was chosen in the software R (R Development Core Team, 2009). Details about the methodology implemented are presented in Appendix B. In the I-T approach inferences can be made from more than one model, something that cannot be done using the traditional model selection approach or the null hypothesis approaches (Johnson and Omland, 2004). The second-order Akaike's information criterion corrected for small sample size (AIC, Hurvich and Tsai, 1989) was used in this research for model selection. The relative probability of each model being the best model was calculated considering their Akaike weights (wi).

When no single model is overwhelmingly supported by the data (i.e. wi max – wi rank) then a model-averaging approach can be used (Gibson et al., 2004). This situation occurs because a number of models in the set may only slightly differ in their data fit, as defined by an information criterion. The advantage of the I-T model-averaging procedure is that it accounts for model selection uncertainty to obtain robust variable estimates or predictions (Grueneberg et al., 2011). To assess the predictive performances in the LRM three criteria were evaluated: 1) percentage of Correctly Classified Observations (Zuur et al., 2011). To assess the predictive performances in the LRM and NBRM. Daily average predictions of these input variables at each sampling station were considered. We consider that using daily average data of water quality and quantity variables at each sampling station as input for the ecological models in the scenarios is a valid approach, because aquatic macroinvertebrates have relatively long life cycles and are confined for most part of their life to one locality on the river bed. Macro-invertebrates integrate water quality over longer periods of time (weeks, months, small years) (Goethals, 2005).

The scenarios were developed for the year 2005 as a reference situation and the year 2015 as projected time period. The year 2005 was considered as reference situation because the Environmental Authority in the Cauca Region (CVC) started a sanitation program in that year and they wanted to evaluate the impact of the program after 10 years (year 2015). The sanitation program plans pollution control measures in the Cauca river basin and includes investments in wastewater treatment plants and clean technologies (CVC and Univalle, 2007). The reference situation (i.e. scenario for year 2005) considered low flow conditions (i.e. flow < 180 m³/s in the Juancho station) and it had detailed information about pollution loads and water quality of the Cauca river in the year 2005. For Juancho sampling station, the CVC and Univalle (2007) reported a range of flows for dry (<180 m³/s), average (180–292 m³/s) and wet conditions (≥292 m³/s). In the framework of the CRMP project a total of 27 scenarios were run and the three most representative scenarios (with 2015 as projected time period) were selected for this study. A projected time period was used to consider the impact of the increase of the wastewater pollution load, due to the growth of the population and the industrial activity in the study area. The four scenarios (i.e. reference situation and three projected scenarios) considered the same river flow characteristics, which means that all considered dry season conditions (low flow), when critical conditions for the dilution of the pollution are observed. Thus, the change in the physicochemical variables (e.g. DO and BOD5) was only related with the pollution control measures proposed in each scenario and was not influenced by a change in the dilution capacity of the river. Projections of average pollution load produced, discharged and removed were calculated (ton/day of BOD5) for each scenario in the study area (Table 2). Moreover, the effective removal percentage, calculated as the ratio between the removed and produced pollution load was reported for each scenario (CVC and Univalle, 2007). The slightly higher value of the effective removal percentage of the scenario of no investment (scenario 2) compared with the current situation (scenario 1), is related with the increase of the amount of pollution load which is removed in the wastewater treatment plant (WWTP). The pollution load removed increased from 2015, 183.6 ton/day of BOD5 (scenario 1) until 256.6 ton/day of BOD5 (scenario 2). This rise is related with the increase of wastewater due to the population growth in districts where there was already a sewer system connected to the WWTP.

3. Results

3.1. Water quality assessment and river water quality modelling

The ecological assessment of the Cauca river showed that BMWP-Colombia values were concentrated only in three EQW classes: class 3 = moderate EQW (moderately polluted); class 4 = deficient EQW (polluted) and class 5 = bad EQW (heavily polluted).

Table 2 Description of the four different pollution control scenarios considered in this research (BODs; five day biological oxygen demand).

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Year</th>
<th>Projection of the average pollution load in the study area (ton/day of BODs)</th>
<th>Effective removal percentage in the scenario − R/P (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Name</td>
<td>P: R</td>
<td>Produced Removed Discharged</td>
<td></td>
</tr>
<tr>
<td>1 Current situation</td>
<td>2005</td>
<td>387.6</td>
<td>183.6</td>
</tr>
<tr>
<td>2 No investment</td>
<td>2015</td>
<td>511.3</td>
<td>256.6</td>
</tr>
<tr>
<td>3 Intermediate situation</td>
<td>2015</td>
<td>511.3</td>
<td>339.5</td>
</tr>
<tr>
<td>4 High investment</td>
<td>2015</td>
<td>511.3</td>
<td>404.9</td>
</tr>
</tbody>
</table>
The sensitivity analysis allowed identifying the most important calibration variables in the MIKE 11 model to predict BOD\textsubscript{5} and DO: the re-aeration rate ($k_2$), the carbonaceous organic matter degradation rate ($k_1$) and the sediment oxygen demand (SOD). The Monte Carlo analysis performed for the calibration process and uncertainty assessment was focused on these three variables. An example of the results of the calibration process of the model for DO and BOD\textsubscript{5} at a specific sampling station (Juanchito) considering dynamic conditions can be seen in Fig. 3a and b. The GLUE technique allowed generating confidence bands for the model results, the higher the confidence band, the higher the uncertainty of the model results. The model performance indicator MSE obtained during the calibration of DO and BOD\textsubscript{5} indicates that for the monitoring stations Hormiguero, Juanchito and Mediacanoa the minimum MSE value was 0.4, whereas for the rest of the stations, Puerto Isaacs and Paso de La Torre, the minimum MSE values were 0.9 and 0.8 respectively.

3.2. River habitat suitability and ecological assessment models

Regarding the collinearity analysis, the first two principal components (PCs) explained 83\% (Spearman correlation coefficient) of the variance in the data. The first PC included temperature, flow, water depth and water velocity, whereas the second PC included BOD\textsubscript{5} and DO. Variables such as BOD\textsubscript{5} and DO ($S = −0.71$), temperature and water velocity ($S = −0.76$), flow and water depth ($S = 0.62$) and flow and water velocity ($S = 0.42$) were highly correlated. DO ($S = −0.76$) and BOD\textsubscript{5} ($S = −0.54$) had the highest correlation with the BMWP-Colombia. In order to avoid highly correlated variables and model overfitting, only DO, water velocity and water depth were kept as predictor variables.

The AIC\textsubscript{c} values, Akaike weights model rankings and performance criteria for all the LRM and NBRM are shown in Table 3. In this table the LRM and NBRM considered are ranked according to their AIC\textsubscript{c} differences ($\Delta_i$), from best to worst. The analysis of the set of “best models” showed that for Ephemeroptera predictions, the first five LRM had $\Delta_i$ lower than four units and good model performances ($CCI > 0.7$, $K > 0.4$ and $AUC > 0.7$). This set of “best models” represents the 95\% confidence set of models-CSR (see cumulative Akaike weights in Table 3). For Haplotaxida predictions, the set of “best models” was confirmed by the first three LRM, with good model performances and represented the 85\% CSM. This indicates that these LRM correctly discriminate between occupied (presence) and unoccupied (absence) sites of these two macroinvertebrate taxa in the dataset. For the BMWP-Colombia predictions, the first six NBRM had $\Delta_i$ lower than four units, however, some of these models had very low performances compared with the others (third and fifth NBRM in Table 3). Therefore, it was decided to eliminate these two NBRM from the set of “best models”, leading to a set of four “best models” with a range of moderate model performance ($r = 0.61−0.69$ and $R^2 = 0.38−0.48$), which represents the 80\% CSM.

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

![Figure 3](https://example.com/figure3.png)

**Figure 3.** Results of the calibration of the Cauca river water quality model at the station Juanchito for (a) dissolved oxygen (DO) and (b) biological oxygen demand (BOD). Simulation period: 22–26 February 2005. Condition: High flows.
Table 4. Additionally, the relative importance of each predictor variable in the 95% confidence set of models is presented in this table. DO and water depth were the most important predictors for Ephemeroptera and the BMWP-Colombia, whereas DO was the most important for Haplotaxida.

The results of the post hoc evaluation of the model adequacy based on diagnostic plots and the lack-of-fit test in the validation of the LRM and NBRM, are presented in the Appendix C1—C3. As an example of this analysis, the most important types of residuals defined in the GLM, the Deviance and the Pearson residuals are presented for the most parsimonious model (lowest AICc). Neither outliers nor high-leverage points nor influential observations were identified. The dispersion parameter (\( \phi \)) for the Poisson regression model in the most parsimonious model was eight. The second alternative (NBRM) did not show any trend in the residual plots and was therefore selected to predict the BMWP-Colombia. The results of the sensitivity analysis of the ecological models are presented in Appendix D. These results confirm those obtained with the I-T approach and showed that DO and water depth were the most important input variables (highest condition number) for the prediction of Ephemeroptera and the BMWP-Colombia, whereas DO was the most important input variable for the prediction of Haplotaxida.

### 3.3. Integrated ecological modelling and scenario assessment

Profiles of average concentrations of DO and BOD\(_5\) at the Cauca river were made for each pollution control scenario considering the results obtained with the MIKE 11 model (Fig. 4). Additionally, the impact of the different scenarios on the EWQ, expressed as the presence/absence of the two target species of macroinvertebrates and the value of the BMWP-Colombia index, was evaluated (Table 5 and Fig. 5a). Furthermore, the EKBI developed in this research and the DO-Prati index were applied on these scenarios (Fig. 5b and c).

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

### 4. Discussion

#### 4.1. Habitat preference and ecological water quality

In the context of the Cauca river management, ecological assessments tools are needed to provide decision makers with accurate information about the EQW, eventually to ensure habitat and species preservation. In order to manage conservation and to restore the river, it is necessary to find out the relationship between the water quality of the river (e.g. physicochemical and hydraulic variables) and the inhabiting organisms (Holguin and Goethals, 2010; Holguin et al., 2013). Models able to predict habitat requirements of organisms may help to insure that planned actions reach the desired effects for the ecosystems (Ahmadi-Nedushan et al., 2006).
The occurrence of Ephemeroptera (i.e. Leptohyphidae and Lep- tophlebiidae families) in the Cauca river was mainly determined by DO and water depth, whereas for Haplotaxida the probability of occurrence varied with DO (i.e. highest relative importance in the 95% confidence set of models; Table 4). The occurrence of these two families of Ephemeroptera was positively related with DO and negatively related with water depth (Table 3), suggesting that these two families are more likely to be found in shallow sites of the Cauca river with high DO concentrations.

Lock and Goethals (2011) stated that most of the species of Ephemeroptera, including those reported for the Leptohyphidae and Leptophlebiidae families, are only present at high DO concentrations and low conductivities. According to Dominguez et al. (2011a) and Lock and Goethals (in press), Ephemeroptera are characteristic for river sites with low human impact, having high DO and low BOD₅ concentrations. In the case of the Cauca river, high DO values indicate good physicochemical water quality, which supports that this taxon is pollution sensitive. It was also found that the presence of Haplotaxida was associated with low DO concentrations (Table 3), suggesting that this species can be present at river sites with high human impact, supporting the concept for this taxon as pollution tolerant.

The occurrence of Ephemeroptera (i.e. Leptohyphidae and Leptophlebiidae families) in the Cauca river was mainly determined by DO and water depth, whereas for Haplotaxida the probability of occurrence varied with DO (i.e. highest relative importance in the 95% confidence set of models; Table 4). The occurrence of these two families of Ephemeroptera was positively related with DO and negatively related with water depth (Table 3), suggesting that these two families are more likely to be found in shallow sites of the Cauca river with high DO concentrations.

Lock and Goethals (2011) stated that most of the species of Ephemeroptera, including those reported for the Leptohyphidae and Leptophlebiidae families, are only present at high DO concentrations and low conductivities. According to Dominguez et al. (2011a) and Lock and Goethals (in press), Ephemeroptera are characteristic for river sites with low human impact, having high DO and low BOD₅ concentrations. In the case of the Cauca river, high DO values indicate good physicochemical water quality, which supports that this taxon is pollution sensitive. It was also found that the presence of Haplotaxida was associated with low DO concentrations (Table 3), suggesting that this species can be present at river sites with high human impact, supporting the concept for this taxon as pollution tolerant.

Regarding the BMWP-Colombia, the most important predictors were DO and water depth (Table 4). This index was positively related with DO and water velocity and negatively with water depth (Table 3). Similar results were reported by Dominguez et al. (2011b) who applied the same index in rivers in Ecuador, and reported that the index scored higher with increasing DO concentrations and high water velocities.

The water quality assessment of the Cauca river showed that the physicochemical indices over predicted the water quality classes.
The predictions of occurrence of Ephemeroptera and Haplo
taxida were determined accurately since the CCI, K and AUC for
the averaged models met the criteria for a good model perform-
ance. The predictions of the BMWP-Colombia index were less
accurate, with 48% of the variance (R²) in the data being explained
by the averaged model, mostly due to the variability that is inher-
ently related to ecological data (Møller and Jennions, 2002;
Symonds and Moussalli, 2011). Ecological models are simplified
representations of the reality, thus, they can never fully predict
nature and always contain errors in assumption, formulation and
parameterization (Lek, 2007; Wannink, et al., 2010). Therefore,
uncertainty assessment of model simulations is important when
models are used to support water management decisions (Beven
and Binley, 1992; Refsgaard et al., 2007).

In general, the results of the calibration and verification process
of the MIKE 11 model, showed that the model was able to accu-
rately predict the dynamic tendencies and the maximum and
minimum values of DO, BOD₅, temperature, flow, water depth and
water velocity for the sampling stations of the Cauca river. The
uncertainty assessment based on the GLUE technique, showed that
the model results were mainly in the range of the 95% confidence
bands, which indicates a good prediction capacity of the model.
These bands allowed quantifying the reliability of the predictions
and represent the influence of the uncertainty related with the
values of the calibration parameters in each monitoring station of
the river.

In this research the multi-model inference method based on
the I-T approach (Burnham and Anderson, 2002) was used as equiva-
 lent to the multiple model simulation described by Refsgaard et al.
(2007). This method allowed us to select a set of “best models”
(using the AICc and the goodness of fit) considering selection un-
certainty. Specific percentages of the confident set of models for the
“best models”; 95% and 85% for the LRM for Ephemeroptera and
Haplotaxida respectively and 80% and 75% for the NBRM were estimated.
As such, we were 95%, 85% and 80% confident that one of the models
within this credibility set is the best approximating model. Addi-
tionally, full multimodel inference was estimated, such as full
model averaged predictions, considering the 95% of confident set
of models. Model-averaged predictions are useful in contexts such
as the one presented here, where there is reasonably high model
uncertainty (i.e. the best AICc model is not strongly weighted),
because predictions are not conditional on a single model (Burnham
and Anderson, 2002). Model averaging recognises that there are two
forms of uncertainty in modelling, the parameter uncertainty and
the model uncertainty. The uncertainty in parameter estimates is
measured by standard errors and confidence intervals for paramete-
rs. Model uncertainty considers that usually the ‘true’ model is un-
known, and there is a probability that each candidate model is the
‘true’ model (Freckleton, 2011). When model uncertainty is present
the I-T approach has considerable advantages over more traditional
step-wise and null-hypothesis approaches to model selection,
where we only end up with a single best model. Model averaged
predictions are likely to be more robust than those derived from a
single best model (Zuur et al., 2009). Moreover, keeping all the
models from the best set of models, allowed us to pick a specific
model with specific predictor variables based on considerations
other than the statistical one, such as the ecological relevance of the
predictors or the model applicability.

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

4.3. Implementation of pollution control scenarios

Considering that the optimal balance between the different
stakeholder activities needs an in depth insight in the integrated
water resources management (Molle, 2009), it is vital that stake-
holders participate in the modelling process (Voinov and Bousquet,
2010). Therefore, in this research four different scenarios for
pollution control in the Cauca river basin were proposed by envi-
ronmental authorities, municipalities and industries. In general, the
scenarios showed that in spite of the reduction of the pollution
load, the DO concentrations in the station Paso de La Torre (abscissa
170.8 km) for all proposed scenarios never reached values for DO
higher than 2.6 mg/L (Fig. 4). Additionally, these DO values are still
lower than the minimum standard value established by the
Colombian legislation (i.e. Decree of 1594 and 1984) for different
uses of the water resource, which means, lower than 70% of the DO
saturation concentration (5.2 mg/L O₂ for this river). The stretch
located between the station Paso de La Torre and Mediacanoa
(above 220.9 km) is the most critical in terms of pollution, mainly
because of the discharge of wastewater coming from the cities of
Cali, Yumbo and Palmira. The habitat suitability models in these
scenarios clearly indicated an improvement in potential habitat
availability for the Ephemeroptera and a decrease in potential
habitat for the Haplotaxida as the pollution load from domestic and
industrial wastewaters is reduced.

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

4.4. Evaluation of the integrated ecological modelling framework

Nowadays, river quality assessment in Colombia relies mainly
on physicochemical standards, however, there is a gap concerning

dated to the multiple model simulation described by Refsgaard et al.
(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.
the impact of different pressures on river biota, which are used to assess river water quality. Some of these pressures are physico-chemical pollution, physical changes and anthropogenic manipulation of the aquatic habitat. The availability and use of decision support tools for water management, such as the one presented in this study, gives an assessment of the impact of these pressures on river biota. By providing an integrated ecological modelling approach, we encourage the integration of different models, data and information resources. Our integrated approach serves, besides its function as a decision support tool, as a communication tool for providing information to the river managers.

Considering the technical point of view, there are two approaches that can be implemented during the integration of models, the integrative approach and the combinatorial approach. The first approach has the benefit of controlling the model design and linkage, but requires longer development time. The second approach saves on the development time, but requires additional work to link up existing models (Lam et al., 2004). However, when a lot of models are already available, the latter is probably the best option. In this research, an intermediate approach that included an existing model for the hydrodynamic and physicochemical components (MIKE 11; CVC and Univaille, 2007) was used and new models (i.e. LRM and NBRM) for the ecological components were developed. This flexible integrated modelling framework allows updating or replacing these regression models by better models when available, without having to change the framework.

In the model development phase different combinations between physicochemical variables (i.e. temperature, BOD$_5$ and DO) and hydro-morphological variables (i.e. flow, water depth and water velocity) were considered. However, there are impacts such as nutrients (i.e. nitrogen and phosphorous), conductivity, particulate inorganic and organic matter, type of bank structure, type of substrate, water body slope and water body sinuosity that may influence the ecological status of rivers (Everaert et al., 2013). Therefore, in order to have a broad spectrum of the EQW and to be able to construct more reliable models, more data should be collected in surface waters characterized by a very good or good EQW and more physicochemical and hydro-morphological variables need to be monitored. Thus, the MIKE 11 model could be used to simulate other processes and to predict some additional variables so that these can be included in the ecological models.

5. Conclusion

In this paper, an ecological modelling framework that integrates a hydraulic and physicochemical water quality model with aquatic ecological models was presented, tested and validated. This generic modelling framework can be used for decision support in river management and water policy as it allows simulation analysis. The technical advance of the integrated modelling framework presented dealt with: 1) simultaneous assessment of hydro-morphological pressures and physicochemical pollution on the EQW; 2) use of a detailed physical habitat and water quality model; 3) development of ecological models based on specific characteristics of the studied river; 4) flexibility for updating or replacing these ecological models by better models when available, without having to change the framework. The integrated framework provides an overview of the effectiveness of the wastewater treatment/disposal strategies and the determination of water quality requirements, considering the receiving water’s ecological aspects. Ecological water quality assessment and habitat suitability models are useful tools for predicting changes in river networks due to disturbances or restoration efforts. These models are a fast and effective way to predict EQW deteriorations or improvements in river systems and allow users to deduce information about a river system that is sometimes unfeasible and very time consuming to monitor. The application of the integrated ecological modelling framework in the Cauca river (Colombia) showed that the currently foreseen investments in sanitation infrastructure will lead to modest improvements of the EQW. Therefore, further actions should be considered to achieve a good EQW.

Acknowledgement

We would like to thank the Environmental Technical Direction of the Environmental Authority in the Cauca Region (CVC), the Natural Resources and Environmental Engineering School of the Universidad del Valle (EIDENAR-Univalle), the Cauca River Modelling Project (CRMP) and Cinara Institute in Colombia, for the information, which was used to develop the research presented in this document. J.E. Holguín-Gonzalez was supported by a doctoral fellowship from the European Union Alfa Project and the Special Research Fund of Ghent University (BOF01/WI0/611 and BOF09/24J/092) in Belgium.

Appendix A. Supplementary material

Supplementary material associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.envsoft.2013.06.004.

References


