Ecological Informatics 7 (2012) 52–61

Contents lists available at SciVerse ScienceDirect

Ecological Informatics

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

Analysis of environmental factors determining the abundance and diversity of macroinvertebrate taxa in natural wetlands of Southwest Ethiopia

Seid Tiku Mereta [a,b], Pieter Boets [a,*], Argaw Ambelu Bayih [b], Asgdom Malu [b], Zewdu Ephrem [b], Addisu Sisay [b], Hailu Endale [b], Menberu Yitbarek [b], Amana Jemal [b], Luc De Meester [c], Peter L.M. Goethals [a]

[a] Laboratory of Environmental Toxicology and Aquatic Ecology, Ghent University, J. Plateaustraat 22, B-9000 Ghent, Belgium
[b] Department of Environmental Health Science and Technology, Jimma University, P.O. Box 378, Jimma, Ethiopia
[c] Laboratory of Aquatic Ecology and Evolutionary Biology, K.U.L., Ch. Deberiotstraat 32, B-3000 Leuven, Belgium

ARTICLE INFO

Article history:
Received 17 June 2011
Received in revised form 7 November 2011
Accepted 20 November 2011
Available online 26 November 2011

Keywords:
Conservation biology
Decision trees
Macroinvertebrates
Sensitivity analysis
Wetlands

ABSTRACT

In Ethiopia, wetland resources play a vital role in the lives of adjacent communities by helping them to achieve food security and livelihoods. However, many wetlands throughout the country are facing degradation as high population growth rate increases the need for more fertile agricultural land. Lack of awareness and logistic constraints are important reasons for the weak consideration of wetland ecosystems by the country's development planners. In this paper, we set out to develop methods for predicting species–environment relationships. Decision tree models and Canonical Correspondence Analysis (CCA) were used to identify factors influencing macroinvertebrate community structure in natural wetlands of Southwest Ethiopia. The models were based on a dataset of 109 samples collected from 57 sites located in eight different wetlands. Sixteen macroinvertebrate taxa were selected based on their frequency of occurrence to determine the status of the wetlands. It was found that Corixidae, Baetidae and Hydrophilidae had the highest predictive model performance. This indicates that these taxa have clear requirements regarding their environmental conditions. The low Kappa value combined with the high number of Correctly Classified Instances of Chironomidae may be related to their high frequency of occurrence, so that their presence is of little predictive power. This was also further illustrated by the Canonical Correspondence Analysis (CCA) where the family of Chironomidae, common at nearly every sampling station in the wetlands, was plotted in the centre of the CCA axis. Vegetation cover, water depth, and conductivity were the most important variables determining the presence or absence of macroinvertebrate taxa. These variables were selected in more than 80% of the classification tree models and played a critical role in the ordination analyses. The sensitivity analysis, based on the regression tree models, also showed that vegetation cover and conductivity were affecting the abundance of some macroinvertebrate taxa. Information on habitat quality and environmental factors preserving a high diversity are essential to develop conservation and management programs for wetlands and their related ecosystem services in Ethiopia, where wetland resources are being lost at a high rate, and continue to be at high risk due to expansion of agricultural and other development activities.

1. Introduction

Wetlands are one of the most biologically productive natural ecosystems on earth (Dixion and Wood, 2003; Rolon and Maltchik, 2006). While they occupy about 6% of the world's land surface, they contribute up to 40% of the annual globe's ecosystem services (Bonell et al., 1993; Costanza et al., 1997). Wetlands perform a wide variety of ecological functions including nutrient cycling (Bunn et al., 1999), carbon storage (Adhikari and Bajracharaya, 2009), flood reduction (Hey and Philippi, 1995) and provisioning of habitat for wildlife (Jacobs et al., 2009). Moreover, wetlands play a vital role in ensuring water supply, food security and livelihoods for millions of people living in developing countries (Shewaye, 2008; Teferi et al., 2010). Understanding the economic value of nature and the services it provides to humanity has become increasingly important for local, national, and global policy and decision making. Quantifying and integrating ecosystem services into decision making will be crucial for sustainable development (Turner et al., 2010). Costanza et al. (1997) calculated that wetland values can contribute worldwide up to $15,000 per hectare per year. The majority of the value of services are currently outside the market system, but should be included in the future. Several of these services include gas regulation, disturbance regulation, waste treatment and nutrient cycling (Costanza et al., 1997). The value of these regulating services derives from the benefits they protect and are, as described by the Millennium Assessment, often related to water quality (Simonit and Perrings, 2011). With

* Corresponding author. Tel.: +32 472521819.
E-mail address: pieter.boets@ugent.be (P. Boets).

1574-9541/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecoinf.2011.11.005

respect to water quality and drinking water production, potential costs in wastewater treatment can be reduced, because tertiary treatment by wetlands may save costs for alternative treatment (Costanza et al., 1997). The threats posed by climate change and the increase in global population predicted to reach nine billion by 2050, resulting in increasing pressures on water resources, urge the need to maximise these benefits (Ramsar, 2010).

Despite the fact that many scientific studies highlight the importance of wetlands for ecosystem services, most wetlands worldwide have suffered from extensive exploitation in the past century (Xu et al., 2011). Studies have shown that about 50% of the world's wetlands have disappeared in the last century due to agriculture and urban development (Mitsch and Gosselink, 1993; Shine and Klemm, 1999). Drainage for agriculture has been recognized as the primary cause of global wetland loss (Xu et al., 2011). In Ethiopia, rapid population growth triggers expansion of agricultural areas, resettlement of landless people, and exploitation activities in wetland areas (Shewaye, 2008). Consequently, several wetlands either disappeared or are on the verge of drying out (Shewaye, 2008), while others rapidly decline in water quality. In response to the rapid degradation of wetlands in Ethiopia, a number of studies on wetland hydrology (Dixion, 2002; Dixion and Wood, 2003) and socio-economic aspects (Solomon, 2004) have been initiated. However, little is known about the overall ecological condition of wetlands in Ethiopia. The diversity and abundance of macroinvertebrates are known to provide considerable information on ecosystem impairment (Feio et al., 2007; Liston et al., 2008). Analysing the health and diversity of these wetlands, based on the presence of macroinvertebrates, could therefore indicate the state of the ecosystem and the related services (Feld et al., 2010). In the present study, we therefore set out to identify the major environmental factors governing the macroinvertebrate communities inhabiting wetlands in a region in Ethiopia that is relatively rich in wetlands, but is under severe pressure by rapidly increasing land use intensity.

Macroinvertebrates represent a diverse group of long living sedentary species that react strongly and often predictably to human influences on aquatic systems (Cairns and Prall, 1993). They are considered very appropriate subjects for the assessment of the ecological condition of wetlands, since they are abundant, readily surveyed, and taxonomically rich (Dodson, 2001). Furthermore, they play an important role in the overall functioning of wetland ecosystems as they occupy a central position in the food web of wetlands (Batzer et al., 1999). Macroinvertebrate community characteristics can reflect primary production and the ability of a wetland to support vertebrate wildlife (e.g. fish) and remove pollutants (Batzer et al., 2006). A better understanding of the factors driving changes in macroinvertebrate community structure along perturbation gradients at several taxonomic levels is therefore important to predict the potential changes in the ecological conditions of wetlands (Trigal-Domínguez et al., 2009).

In order to predict the habitat requirements of wetland macroinvertebrate communities, there is a clear need for models quantifying species–environment relationships to support decision making (Broekhoven et al., 2006). Modelling the distribution of taxa as a function of the abiotic environment, often called habitat suitability modelling, has been recognized as a significant component of conservation planning (Guisan and Zimmermann, 2000). Habitat suitability models combine occurrence and/or abundance of species with environmental variables, both biotic and abiotic factors, judging on the habitat quality or predicting the effect on species occurrence of environmental changes within the habitat (Anderson et al., 2003; Store and Kangas, 2001). These models are typically developed by identifying statistical relationships between the occurrence and/or the abundance of the species and the biochemical and physical properties of a given site (Store and Kangas, 2001). In this regard, many approaches including multivariate analysis (Robertson et al., 2001) and modelling techniques such as decision trees (Boets et al., 2010; Dakou et al., 2007; Goethals et al., 2002; Hoang et al., 2010), artificial neural networks (Dedecker et al., 2007; Goethals et al., 2007; Park et al., 2003), fuzzy logic (Broekhoven et al., 2006; Mouton et al., 2009) and Bayesian belief networks (Adriaenssens et al., 2004) have been applied.

The aim of the present study was to analyse the relationship between habitat quality and the occurrence and diversity of macroinvertebrates in Wetlands in Southwest Ethiopia. Therefore, we developed habitat suitability models using decision tree models and used multivariate data analysis in order to analyse the macroinvertebrate community structure in these natural wetlands. The information obtained from this study can be used to inform on environmental factors that are important for community structure of macroinvertebrates and as a guideline for habitat conservation of wetlands and their related ecosystem services.

2. Methods and materials

2.1. Study area

The data used for the present study were collected from wetlands located in the Gilgel Gibe watershed, Southwest Ethiopia (Fig. 1). Six permanent (Koffe, Kitto, Boye, Haro, Bulbul and Balawajo) and two temporary (Haro1 and Haro 2) wetlands located along the Gilgel Gibe river were included. The studied wetlands are varying in size ranging from 5 ha to a few hundred hectares. These wetlands serve as a source of drinking water, as breeding grounds for birds and as grazing land (Yimer and Mengistou, 2009). All permanent wetlands except Bulbul are riverine, connected upstream and downstream to the rivers flowing into the Gilgel Gibe River and finally to the Gilgel Gibe hydro power dam. The temporary and Bulbul wetlands are created by a meandering floodplain. These temporary wetlands are characterised by high fish and waterfowl abundance. The major threats from human activities around and in these wetlands include intensive livestock grazing, brick making, vegetation clearance, land conversion to cropland, drainage, municipal waste discharge and cultivation. Maize (*Zea mays*) cultivation is a common practice in and around these wetlands.

2.2. Data collection

A total of 57 sampling stations were monitored. Fifty two permanent sampling sites were sampled both during the dry (March to May, 2010), and the wet (August to September, 2010) season, whereas five temporary wetland sampling stations were sampled only during the wet season. In this way, 109 samples were available.

2.2.1. Abiotic habitat characteristics

Abiotic habitat characteristics at each sampling station over a 500-meter reach were assessed using the USEPA wetland habitat assessment protocol (Baldwin et al., 2005). The physical features measured included vegetation cover, water depth, hydromorphological settings and adjacent land use patterns (grazing, cultivation/ploughing, clay mining, drainage and waste dumping; see Table 1).

Dissolved oxygen, electric conductivity, pH and water temperature were measured in the field using a multi-probe meter (HQ30d Single-Input Multi-Parameter Digital Meter, Hach). Chlorophyll a concentration was measured on site using a fluorometer (Turner Design Aquafluor). At each site 2 l of water was collected and stored on ice until return to the Laboratory of Environmental Health at Jimma University, where samples were analysed for total nitrogen (TN), total phosphorus (TP), five day biochemical oxygen demand (BOD_5), chemical oxygen demand (COD), orthophosphate, ammonium and nitrate concentration according to the standard methods as prescribed by APHA et al. (1995).

2.2.2. Biotic habitat characteristics

Macroinvertebrates were collected at each sampling station using a rectangular frame net (20 × 30 cm) with a mesh size of 300 μm. Each

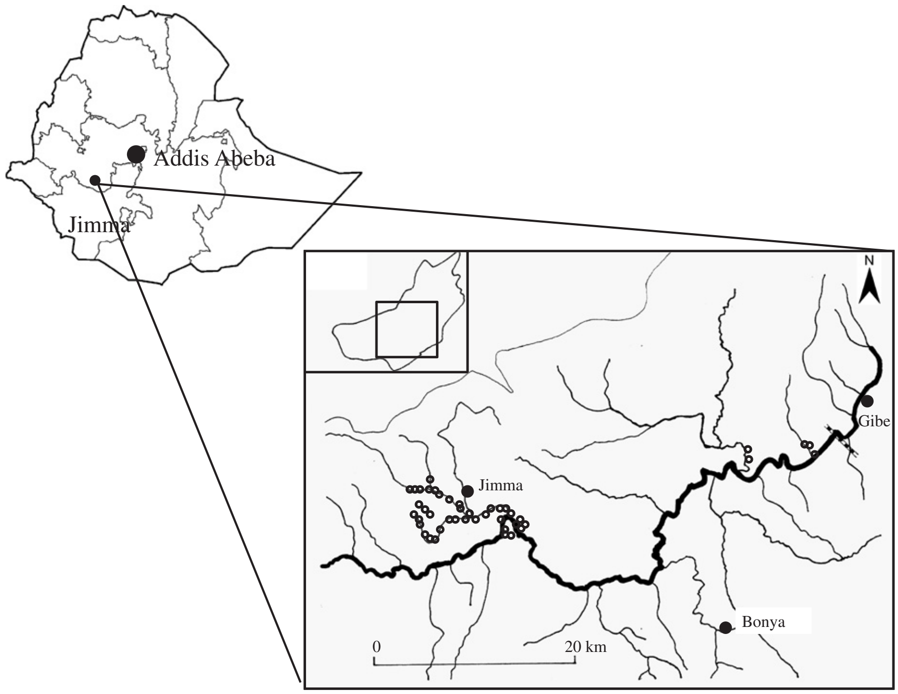

Fig. 1. Location of the study area and wetland sampling stations (black circles) in the Gilgel Gibe watershed, Southwest Ethiopia.

collection entailed a 10-minute kick sampling over a distance of 10 m (DNRE, 1999). Time was allotted proportionally to the cover of different meso-habitats of the wetland such as open water and emergent vegetation. The bottom sediment was disturbed by foot during sampling in order to collect the benthic macroinvertebrates. Macroinvertebrates were sorted in the field, stored into vials containing 80% ethanol and labelled. Afterwards, macroinvertebrates were identified to family level using a stereomicroscope (10× magnification) and the identification key of Bouchard (2004).

Table 1
Input variables used for the model development: mean values, standard deviation, and range. TON = total organic nitrogen, NH_4^+ = ammonium, NO_3^- = nitrate, TP = total phosphorus, PO_4^{3-} = orthophosphate, BOD_5 = biological oxygen demand, COD = chemical oxygen demand.

Variables	Unit	Mean	Standard deviation	Range
Ambient temperature	°C	25	3	17–34
Water temperature	°C	22	2	18–33
pH	–	7	0.6	6–10
Dissolved oxygen	mg/l	4	2.5	0.2–14
Oxygen saturation	%	55	41	2–263
Conductivity	μS/cm	103	55	41–293
Chlorophyll a	μg/l	13	2	11–22
TON	mg/l	5	6	0.05–34
NH_4^+	mg/l	0.10	0.2	0.01–1.6
NO_3^-	mg/l	1.4	1.9	0.04–12
TP	mg/l	0.2	0.2	0.03–1.2
PO_4^{3-}	mg/l	0.15	0.5	0.01–5.4
BOD_5	mg/l	14	17	1–144
COD	mg/l	25	36	3–306
Water depth	Cm	49	27	5–180
Vegetation cover	%	69	18	35–95
Fish	Absent (0), present (1)	N/A	N/A	N/A
Grazing	Absent (0), present (1)	N/A	N/A	N/A
Cultivation/ploughing	Absent (0), present (1)	N/A	N/A	N/A
Clay mining	Absent (0), present (1)	N/A	N/A	N/A
Drainage	Absent (0), present (1)	N/A	N/A	N/A
Waste dumping	Absent (0), present (1)	N/A	N/A	N/A

Each site was sampled for fish during both the dry (March to May, 2010) and wet season (August to September, 2010). Fyke nets were used as well as fish pots. These were positioned in shallow areas (less than 1 m depth) at each site. Nets were set during the day and retrieved after about 24 h. Fish were counted and their length and weight were measured before they were released.

2.3. Data analysis

Multivariate statistical analysis and several classification and regression tree models were used to analyse the habitat preference of macroinvertebrate taxa in the sampled wetlands.

2.3.1. Classification and regression tree models (CART)

Twenty two environmental variables were used to determine the most important variables for the prediction of the 16 most frequently occurring macroinvertebrate taxa in the wetlands (Tables 1 and 2). Classification and regression tree models (CART) were applied to develop the models. The classification tree models were built using the J48 algorithm (Quinlan, 1993), a java re-implementation of the C4.5 algorithm, which is a part of machine learning package WEKA (Witten and Frank, 2005). Regression tree models were built using the M5 algorithm in WEKA (Witten and Frank, 2005) in order to relate the abundance of macroinvertebrate taxa to environmental variables. Default parameter settings were used to induce the trees.

Model training and validation were based on a three-fold cross validation procedure (Witten and Frank, 2005). All data (dry and wet season sampling data) were used to construct the models. The dataset was stratified into three subsets, of which two subsets were used as training data and the remaining one subset was used for testing the model. The cross validation process was then repeated three times each with one of the three subsets used once as the validation data set. In this way, three models were built. The results from the three models were averaged to produce a single prediction of the dependent variable as well as the variation based on the difference between the outcome of the three models.

Table 2
Overview of the identified taxa as well as their frequency of occurrence in all samples.

Family	Frequency of occurrence (%)
Caenidae	23
Simuliidae	23
Tipulidae	28
Culicidae	28
Helodidae	30
Sphaeriidae	36
Planorbidae	42
Baetidae	50
Notonectidae	50
Belostomatidae	58
Corixidae	61
Libellulidae	64
Coenagrionidae	71
Hydrophilidae	72
Dytiscidae	83
Chironomidae	87

The percentage of correctly classified instances (CCI) (Witten and Frank, 2005) and Cohen's Kappa statistic (K) (Cohen, 1960) were used to evaluate the predictive performance of the classification tree models. The CCI is the percentage of the true positive (TP) and true negative (TN) predictions, which is calculated based on a confusion matrix.

CCI is mathematically expressed as follows:

$$CCI = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

Cohen's Kappa statistic simply measures the proportion of all possible cases of presence or absence that are predicted correctly by a model after accounting for chance predictions. It is mathematically expressed as follows:

$$K = \frac{[(TP + TN) - (((TP + FN)(TP + FP) + (FP + TN)(FN + TN))/n)]}{[n - (((TP + FN)(TP + FP) + (FP + TN)(FN + TN))/n]}$$

where n is the total number of instances, TP is the percentage of true positives, TN the percentage of true negatives, FP the percentage of false positives and FN the percentage of false negatives.

Models with a CCI higher than or equal to 70% and K higher than or equal to 0.4 were considered reliable (Dakou et al., 2007; Gabriels et al., 2007). CCI is affected by the frequency of occurrence of the taxon being modelled (Manel et al., 2001). Unlike CCI, K takes a correction into account for the expected number of correct predictions due to randomness, which is strongly related to taxon prevalence (Fielding and Bell, 1997; Manel et al., 2001). We used the ranges of K recommended by Landis and Koch (1977) for model performance evaluation: $K \leq 0$ (poor), 0–0.2 (slight), 0.2–0.4 (fair), 0.4–0.6 (moderate), 0.6–0.8 (substantial) and 0.8–1 (almost perfect).

We used the determination coefficient (R^2) value to evaluate the performance of the regression tree models (De'ath and Fabricius, 2000). The determination coefficient is a measure of the goodness of fit of the model predictions to the training data (Kallimanis et al., 2007). Its value is always between 0 and 1. The closer the value is to 1, the better the model predicts the training data.

Sensitivity analysis was performed in order to gain insight in the relationship between predictor variables and the abundance of macroinvertebrate taxa. For each of the three models constructed per taxon, the gradient and importance of the predictor variable (e.g. conductivity) on the macroinvertebrate abundance were analysed. This was done by plotting the selected variable between its minimum and maximum values encountered at the sampling sites, while the other parameters that were present in the model were kept constant at their average values. In this way, for each of the three different models (folds) a line was plotted showing the relationship between the environmental factors and the abundance of macroinvertebrates as well as the gradient of the different models.

2.3.2. Multivariate data analysis

Detrended Correspondence Analysis (DCA) was applied using CANOCO 4.5 (ter Braak and Smilauer, 2002) to examine whether Redundancy Analysis (RDA) or Canonical Correspondence Analysis (CCA) would be appropriate (ter Braak and Jaap, 1994) to analyse the data. The DCA yielded gradient lengths that were higher than three standard deviations, therefore CCA was used. Sixteen macroinvertebrate taxa were selected based on their frequency of occurrence. Macroinvertebrate abundance data were log transformed log(x + 1) prior to analysis to obtain homogeneity of variance. Based on a stepwise forward selection twelve environmental factors were selected as independent variables. All environmental data except pH and presence of fish were log(x + 1) transformed and standardized since the variables were measured in a variety of units. The statistical significance of eigenvalues and species–environment correlations generated by the CCA were tested using Monte Carlo permutations. All data (dry and wet season sampling data) were used together to construct the plots.

3. Results

3.1. Variable importance

Twenty two environmental variables (Table 1) were used as predictors to determine the presence/absence of 16 benthic macroinvertebrate taxa (Table 2). Fig. 2 shows the average frequency of selection of these environmental variables by the classification tree models as well as the variation on the different models. Since the training and validation were based on three-fold cross validation, three models were developed for each taxon. In total, 48 models (three models per taxon) were constructed.

The most frequently selected variables were vegetation cover (88%), conductivity (81%), water depth (81%), presence/absence of fish (56%), and total phosphorus concentration (56%). Moreover, vegetation cover and conductivity were often selected as root of a tree indicating that these were the most informative attributes to determine the presence/absence of macroinvertebrates taxa. In contrast to the above mentioned variables, ambient and water temperature and chlorophyll a were selected in 6% of the cases and thus were less critical for explaining the presence/absence of taxa.

As an example, the classification tree model for Caenidae is depicted in Fig. 3. This tree has seven leaves and thirteen branches. The classification tree indicates that vegetation cover, given as a root of the tree, is considered as the most informative attribute to predict the occurrence of Caenidae. Caenidae were generally absent when the vegetation cover was less than 55% and pH was higher than 7.23. On the other hand, Caenidae were present in sites where there was no clay mining activity and fish were absent. This classification tree had a good overall predictive performance, with a CCI of 81% and Kappa of 0.47.

3.2. Model performance evaluation

The model performances based on the CCI and Cohen's Kappa statistic of the three-fold cross validation for 16 macroinvertebrate taxa are shown in Fig. 4A and B. Error bars give the variation between the subset models constructed per taxon. The CCI varied between 59 ± 3% and 88 ± 2%. Based on CCI, 13 taxa were predicted with a good reliability by the classification tree models (CCI ≥ 70%). Based on CCI, very good predictions were obtained for Chironomidae and Dytiscidae with CCI of 88 ± 2% and 86 ± 11%, respectively. On the other hand, eight taxa were predicted accurately based on Cohen's Kappa statistic ($K \geq 0.4$). Based on Kappa, the highest model predictive

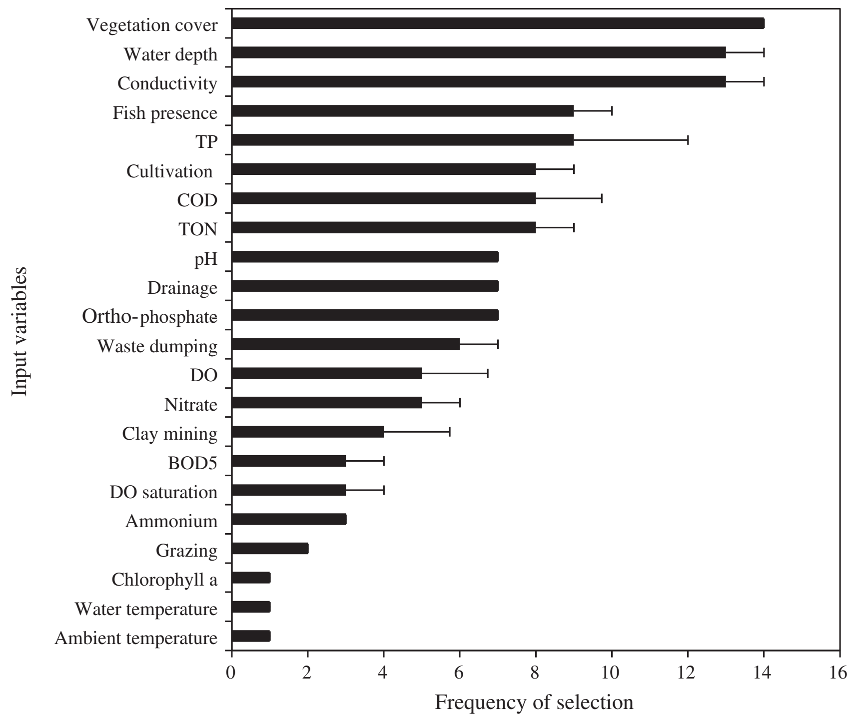

Fig. 2. Overview of the average frequency of selection and standard variation of the different input variables used in the decision tree models to model the presence or absence of each taxon.

performance was obtained for Corixidae and Baetidae with a K value of 0.58 ± 0.2, indicating a good model performance. In contrast, Tipulidae and Belostomatidae had the lowest K value (0.17 ± 0.02), indicating poor model performance.

Although Chironomidae and Dytiscidae had the highest relative CCI, their K values were 0.30 ± 0.27 and 0.38 ± 0.3, respectively. The high CCI values were related to their high frequency of occurrence: the family Chironomidae was present in 87% and Dytiscidae in 83% of the samples (Table 2).

3.3. Sensitivity analysis

We carried out a sensitivity analysis for the six macroinvertebrate taxa with an acceptable model performance, namely Caenidae,

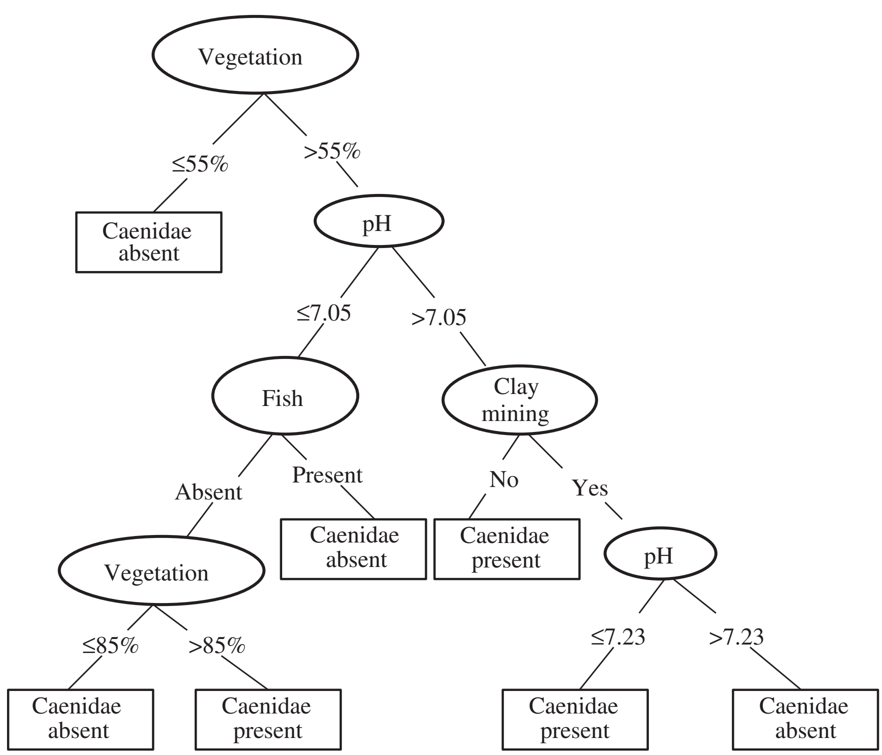

Fig. 3. Classification tree model predicting the presence or absence of Caenidae (Correctly Classified Instances = 81%, Kappa = 0.47).

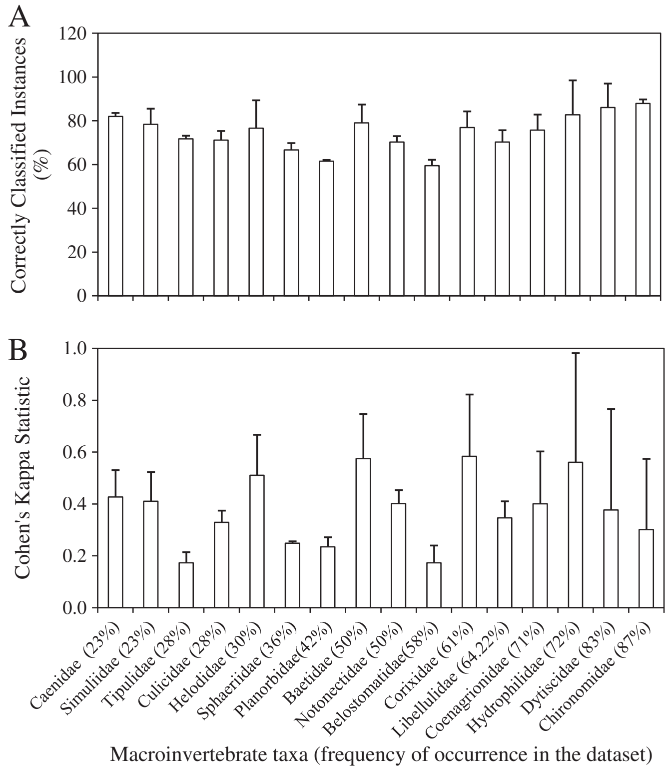

Fig. 4. Overview of the average predictive performance of each model as well as the variation on the models based on (A) Correctly Classified Instances, (B) Cohen's Kappa statistic of all macroinvertebrate taxa that were modelled. The percentage of occurrence of each taxon at the different sampling sites is given between brackets.

Baetidae, Simuliidae, Dytiscidae, Hydrophilidae and Notonectidae. The correlation coefficient obtained from the regression tree models for these six taxa varied from 0.29 ± 0.02 to 0.55 ± 0.12. Vegetation cover and conductivity were used as predictor variables since these were the most important variables selected by the models.

The sensitivity analysis pointed out that the abundance of Simuliidae (Fig. 5A) and Baetidae (Fig. 5B) increased with increasing vegetation cover. Caenidae abundance increased up to 80% vegetation cover and became more or less stable afterwards (Fig. 5C). In contrast to the other taxa, the abundance of Notonectidae decreased with increasing vegetation cover (Fig. 5D).

A sensitivity analysis of the regression tree model analysing the effect of changing conductivity on the abundance of Simuliidae, Dytiscidae and Hydrophilidae is shown in Fig. 6. The abundance of Simuliidae (Fig. 6A) decreased with increasing conductivity. In contrast, the abundance of Coleoptera larvae, both Dytiscidae and Hydrophilidae (Fig. 6B,C), increased with increasing conductivity and remained more or less stable at conductivity levels above 150 μS/cm.

3.4. Multivariate analysis

The first and the second canonical axes explained 13.7% (eigenvalue of 0.17) and 7.8% (eigenvalue of 0.10) of the variation in the species data, respectively. The species–environment correlation of the first axis was statistically significant in a Monte Carlo permutation test ($P<0.05$). The first axis was positively correlated with the presence/absence of fish ($r=0.62$), total phosphorus ($r=0.61$), water depth ($r=0.54$), dissolved oxygen ($r=0.53$), and chemical oxygen demand ($r=0.43$). Vegetation cover and conductivity were negatively correlated with CCA axis 1, with $r=-0.27$ and $r=-0.38$, respectively. CCA axis 2 was positively correlated with vegetation, conductivity, COD and TP, and negatively with dissolved oxygen (Fig. 7). In addition, CCA analysis also revealed that Simuliidae and Caenidae were significantly correlated with vegetation cover ($r=0.5$ and $r=0.42$, respectively; $P<0.05$). Hydrophilidae ($r=0.45$) and Dytiscidae ($r=0.47$) were significantly correlated with water conductivity ($P=0.04$).

A bi-plot of the sampling sites and environmental variables showed that there was a clear distinction between samples taken during the dry or the wet season (Fig. 8). Conductivity was strongly positively correlated with dry season, whereas vegetation cover and

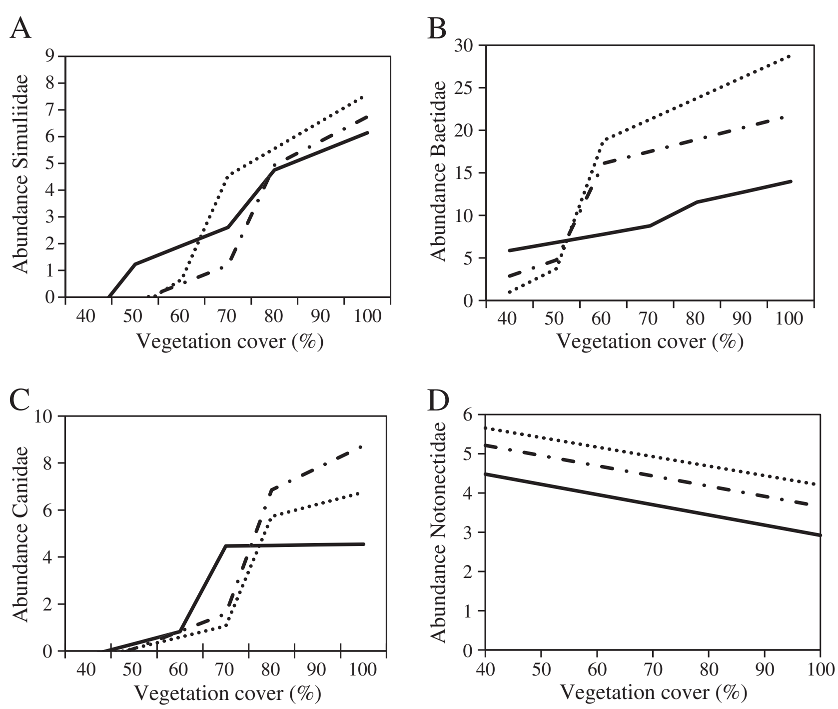

Fig. 5. Sensitivity analysis illustrating the abundance (number of individuals per sample) of (A) Simuliidae, (B) Baetidae, (C) Caenidae, (D) Notonectidae in function of vegetation cover (Fold 1 = dotted line, Fold 2 = solid line, Fold 3 = dashed line; for more explanation on folds, see text).

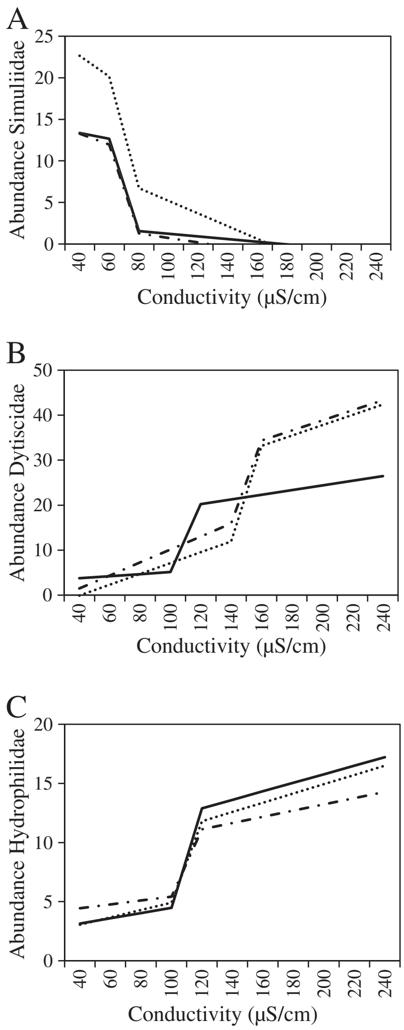

Fig. 6. Sensitivity analysis illustrating the abundance (number of individuals per sample) of (A) Simuliidae (B) Dytiscidae and (C) Hydrophilidae in function of conductivity (Fold 1 = dotted line, Fold 2 = solid line, Fold 3 = dashed line; for more explanation on folds, see text).

dissolved oxygen were more correlated with the wet season samples. Temporary wetlands as well as wetlands with a lot of open water clustered together and showed associations with water depth, COD, TP and the presence/absence of fish.

4. Discussion

Predicting species' distributions has been recognized as a significant component of conservation planning since it helps identifying those regions which yield maximum effect when including restoration efforts (Guisan and Zimmermann, 2000; Loiselle et al., 2003). In the present paper, predictive models allowed identifying important variables structuring the macroinvertebrate community in wetlands in Southwest Ethiopia. Using Kappa values as indicator, our models performed least well for the Tipulidae and Belostomatidae, suggesting that other factors than the ones we quantified determined the distribution of these taxa. The low kappa value and the high CCI of the models for Chironomidae may be related to their high frequency of occurrence, indicating that the predictions could easily be generated by chance (Fielding and Bell, 1997; Manel et al., 2001). The model indeed tends to learn that the most common taxa are always present and the rarest taxa are always absent (Dedecker et al., 2007). In the CCA, Chironomidae were plotted in the centre of the axes. The weak association between Chironomidae and environmental factors and the fact that this species occurred in 87% of the sites reflect that this taxon is tolerant to disturbance and a resident of impacted environments, as has been reported by many earlier studies (Karr and Rossano, 2001). Corixidae, Baetidae and Hydrophilidae showed high Kappa and CCI values, and their occurrence could be well predicted by our model. This indicates that these taxa have clear requirements regarding their environmental conditions within the habitat gradient we studied.

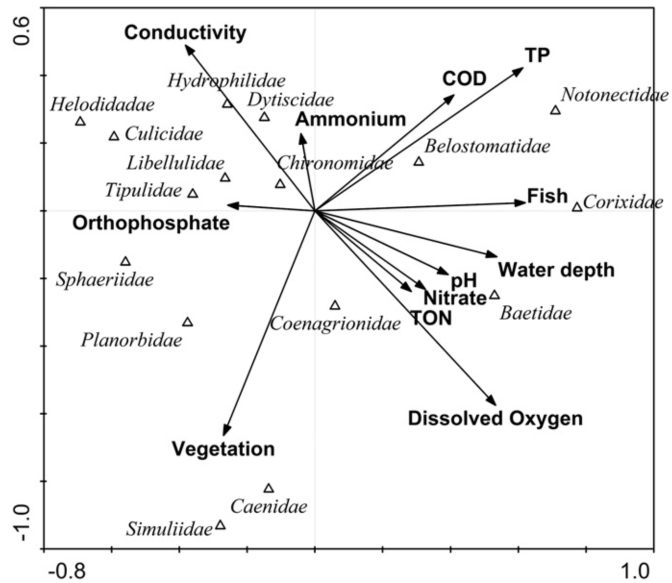

Fig. 7. Canonical correspondence analysis (CCA) of macroinvertebrate taxa and environmental variables in natural wetlands of Southwest Ethiopia (variables are explained in Table 1).

Vegetation cover, water depth and conductivity were the most important environmental variables determining the presence or absence of macroinvertebrate taxa. These variables were selected in more than 80% of the classification tree models and were also correlated with the axes that explain the largest amount of variation in the ordination analysis. Vegetation is known to be an important parameter in wetlands influencing the diversity of macroinvertebrates (Balcombe et al., 2005; Jurado et al., 2009). Both the sensitivity analysis

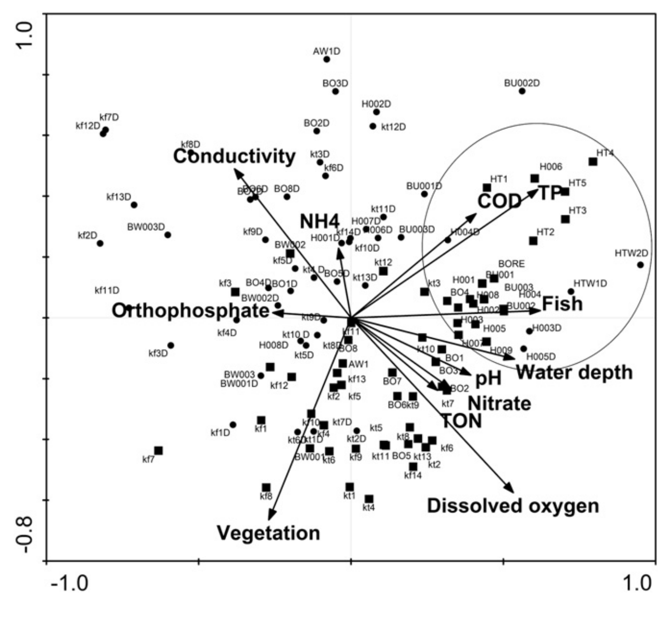

Fig. 8. Bi-plot of environmental variables and wet (squares) and dry season (circles) sampling sites. The temporary wetlands (wet season samples only) are clustered and indicated by a circle (variables are explained in Table 1).

and ordination diagram showed that the abundance of Simuliidae and Caenidae is strongly associated with vegetation cover. Vegetation can provide shelter against water current and predation, can provide more food resources, and is important as oviposition site (Ambelu et al., 2010; Couceiro et al., 2007). Vegetation has been shown to decrease the efficiency of fish predation and provides a refuge for benthic macroinvertebrates against visual predators (Diehl, 1995; Hanson and Butler, 1994). In addition, macrophytes produce dissolved oxygen trough photosynthesis and in this way, create better habitat conditions (Ságová Marecková and Kvet, 2002). In contrast to the other taxa, the abundance of Hemiptera was negatively correlated with vegetation cover. The crucial factor for the distribution of aquatic Hemiptera species seems to be a high percentage of open water, as also shown in other studies (Bloechl et al., 2010).

Besides vegetation, conductivity is an important parameter affecting the composition of macroinvertebrate communities (Boets et al., 2010; Gabriels et al., 2007). Although the measured conductivity values in our study were generally low, ranging between 41 and 293 μS/cm, several wetlands were strongly influenced by inflow of untreated wastewater from Jimma town, which led to an important input of water with a relatively high conductivity. Several studies have shown that urbanization can contribute to increased levels of conductivity in fresh water ecosystems (Roy et al., 2003). Both the ordination and sensitivity analysis showed that taxa belonging to the order of the Coleoptera were positively correlated with conductivity. The preference of Coleoptera for relatively high levels of conductivity can be species dependent, as found by Cuppen (1986). Conductivity was lower in water samples taken during the wet season, likely reflecting dilution effects by runoff and precipitation. Culicidae larvae were positively correlated with conductivity and negatively correlated with dissolved oxygen concentration. Culicidae have a breathing tube siphon that allows them to obtain oxygen from the air and thus persist in environments with poor water quality (Chipps et al., 2006). As most other organism groups cannot cope with low oxygen levels, the Culicidae are released from competitive pressure in low oxygen habitats, which increases their likelihood of occurrence. Dissolved oxygen concentrations were higher in the wet season samples, suggesting that runoff, precipitation and turbulence increased the dissolved oxygen concentration (Ambelu, 2009). Since some variables are strongly negatively correlated, extra analyses on variable selection are necessary to give inconclusive results on which parameter is the most important.

Temporary wetlands and wetlands with a lot of open water differed from permanent wetlands in their macroinvertebrate composition as well as in their physical and chemical characteristics. Several studies have shown that the hydroperiod plays a critical role in the ecology of wetlands (Steinman et al., 2003). Macroinvertebrate assemblages of temporary wetlands are often characterised by rapidly developing and very active species, or by species that have very high dispersal capacities (Wellborn et al., 1996). The ordination analysis revealed that Hemiptera, Notonectidae, Corixidae and Belostomatidae dominated in temporary wetlands. These taxa are able to re-colonize temporary wetlands within a couple of weeks after flooding (Chase and Knight, 2003). Moreover, in seasonal habitats such as wetlands, community structure can be related to abiotic variables that change in response to seasonal conditions, including water depth, dissolved oxygen and macrophyte coverage (Escalera-Vazquez and Zambrano, 2010).

Fish were mainly found in open water of the temporary wetlands. Several studies demonstrated that the density of fish strongly affects the abundance and distribution of Dytiscidae and other macroinvertebrate taxa (Arnott and Jackson, 2006). In our case, we only took into account the presence or absence of fish, not their abundance. Nevertheless, it is expected that the fish community has an impact on the abundance of macroinvertebrates. This might explain why the abundance and diversity of macroinvertebrates were generally lower in the temporary compared to the permanent wetlands.

The high chemical oxygen demand and high concentration of total phosphorus observed in the temporary wetlands are probably due to agricultural waste products and litter decomposition. Most temporary wetlands were situated in areas with intensive agricultural activity. In the study area, the temporary wetlands are often used as agricultural field or grazing land during the dry season. Cattle can deposit significant amounts of excrements in these fields. When these areas become inundated during the rainy season, the dead organic material from crops and cattle excrements can be decomposed and results in an increase of the concentration of total phosphorus and an increase in chemical oxygen demand (Del Rosario et al., 2002; Strand and Merritt, 1999).

In conclusion, both the decision tree models and the canonical correspondence analysis indicated that environmental factors such as vegetation cover, water depth and water conductivity influence the structure of wetland macroinvertebrate communities. These variables gave a clear and stable result that was easy to interpret. One of the prioritizing services wetlands provide are water and food supply, which have become more scarce due to a growing human population (Ramsar, 2010). A minimal preservation of vegetation in these wetlands is essential to maintain a high biodiversity and to protect ecosystem services. Besides the well known ecosystem services wetlands already provide, biodiversity conservation may provide direct benefits through ecotourism. A good management program for wetlands in Ethiopia is important as their social–economic values are numerous and of high importance to people worldwide (Barbier et al., 2009; Costanza et al., 1997). Further study is recommended to elucidate the ecological implications of the environmental factors identified here on waterfowl and fish, since this could contribute to an improved management of the study wetlands.

Acknowledgements

The authors are grateful to VLIR-UOS for financing this study. The first author is a recipient of an IUC (Institutional University Cooperation) PhD scholarship from VLIR to carry out this work. The authors wish to thank the many people who helped with field work, and the IUC Jimma University program management team for logistic support.

References

Adhikari, S., Bajracharaya, S., 2009. A review of carbon dynamics and sequestration in wetlands. Journal of Wetlands Ecology 2, 2–46.

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, 181–191.

Ambelu, B.A., 2009. Biological monitoring based on macroinvertebrates for decision support of water management in Ethiopia. PhD thesis, Ghent University, Ghent Belgium.

Ambelu, A., Lock, K., Goethals, P.L.M., 2010. Comparison of modelling techniques to predict macroinvertebrate community composition in rivers of Ethiopia. Ecological Informatics 5, 147–152.

Anderson, R.P., Lew, D., Peterson, A.T., 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecological Modelling 162, 211–232.

APHA, AWWA, WPCF, 1995. Standard Methods for the Examination of Water and Wastewater, 19th ed. American Public Health Association, Washington D.C.

Arnott, S.E., Jackson, A.B., 2006. Distribution and potential effects of water beetles in lakes recovering from acidification. Journal of the North American Benthological Society 25, 811–824.

Balcombe, C.K., Anderson, J.T., Fortney, R.H., Kordek, W.S., 2005. Aquatic macroinvertebrate assemblages in mitigated and natural wetlands. Hydrobiologia 541, 175–188.

Baldwin, D.S., Nielsen, D.L., Bowen, P.M., Williams, J., 2005. Recommended Methods for Monitoring Floodplains and Wetlands. 1921038 20 9. (MDBC Publication No. 72/04).

Barbier, E.B., Baumgärtner, S., Chopra, K., Costello, C., Duraiappah, A., Hassan, R., Kinzig, A., Lehman, M., Pascual, U., Polasky, S., Perrings, C., 2009. The valuation of ecosystem services. In: Naeem, S., Bunker, D., Hector, A., Loreau, M., Perrings, C. (Eds.), Biodiversity, Ecosystem Functioning, and Human Wellbeing: An Ecological and Economic Perspective. Oxford University Press, Oxford.

Batzer, D.P., Rader, R.B., Wissinger, S.A., 1999. Invertebrates in Freshwater Wetlands of North America: Ecology and Management. John Wiley and Sons, New York.

Batzer, D.P., Cooper, R., Wissinger, S.A., 2006. Wetland animal ecology. In: Batzer, D.P., Sharitz, R.R. (Eds.), Ecology of Freshwater and Estuarine Wetlands. University of California Press, Berkeley, CA, USA, pp. 242–284.
Bloechl, A., Koenemann, S., Philippi, B., Melber, A., 2010. Abundance, diversity and succession of aquatic Coleoptera and Heteroptera in a cluster of artificial ponds in the North German Lowlands. Limnologica 40, 215–225.
Boets, P., Lock, K., Messiaen, M., Goethals, P.L.M., 2010. Combining datadriven methods and lab studies to analyse the ecology of *Dikerogammarus villosus*. Ecological Informatics 5, 133–139.
Bonell, M., Hufschmidt, M.M., Gladwell, J.S., 1993. Hydrology and Water Management in the Humid Tropics. Cambridge university press. doi:10.2277/0521452686.
Bouchard, R.W., 2004. Guide to aquatic macroinvertebrates of the Upper Midwest; Water Resources Center. University of Minnesota, St. Paul, MN (208PP.).
Broekhoven, E.V., Adriaenssens, V., De Baets, B., Verdonschot, P.F.M., 2006. Fuzzy rule-based macroinvertebrate habitat suitability models for running waters. Ecological Modelling 198, 71–84.
Bunn, S.E., Davies, P.M., Mosisch, T.D., 1999. Ecosystem measures of river health and their response to riparian and catchment degradation. Freshwater Biology 41, 333–345.
Cairns, J.R., Prall, J.R., 1993. A history of biological monitoring using benthic macroinvertebrates. In: Rosenberg, D.M., Resh, V.H. (Eds.), Freshwater Biomonitoring and Benthic Macroinvertebrates. Chapman and Hall, New York, pp. 159–194.
Chase, J.M., Knight, T.M., 2003. Drought-induced mosquito outbreaks in wetlands. Ecology Letters 6, 1017–1024.
Chipps, S.R., Hubbard, D.E., Werlin, K.B., Haugerud, N.J., Powell, K.A., Thompson, J., Johnson, T., 2006. Association between wetland disturbance and biological attributes in floodplain wetlands. Wetlands 26, 497–508.
Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20, 37–46.
Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neil, R.V., Paruelo, J., Raskin, R.G., Sutton, P., van den Belt, M., 1997. The value of the world's ecosystem services and natural capital. Nature 387, 253–259.
Couceiro, S.R.M., Hamada, N., Luz, S.L.B., Forsberg, R.B., Pimentel, T.P., 2007. Deforestation and sewage effects on aquatic macroinvertebrates in urban streams in Manaus, Amazonas, Brazil. Hydrobiologia 575, 271–284.
Cuppen, J.G.M., 1986. On the habitats, distribution and life-cycles of the Western European species of the genus *Helochares Mulsant* (Coleoptera: Hydrophilidae). Hydrobiologia 132, 169–183.
Dakou, E., D'heygere, T., Dedecker, A.P., Goethals, P.L.M., Lazaridou-Dimitriadou, M., De Pauw, N., 2007. Decision tree models for prediction of macroinvertebrate taxa in the river Axios (Northern Greece). Aquatic Ecology 41, 399–411.
De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81, 3178–3192.
Dedecker, A., Van Melckebeke, K., Goethals, P.L.M., De Pauw, N., 2007. Development of migration models for macroinvertebrates in the Zwalm river basin (Flanders, Belgium) as tools for restoration management. Ecological Modelling 203, 72–86.
Del Rosario, R.B., Betts, E.A., Resh, V.H., 2002. Cow manure in headwater streams: tracing aquatic insect responses to organic enrichment. Journal of the North American Benthological Society 21, 278–289.
Diehl, S., 1995. Direct and indirect effects of omnivory in a littoral lake community. Ecology 76, 1727–1740.
Dixion, A.B., 2002. The hydrological impacts and sustainability of wetlands drainage cultivation in Illubabour Ethiopia. Land Degradation and Development 13, 17–31.
Dixion, A.B., Wood, A.P., 2003. Wetland cultivation and hydrological management in Eastern Africa: matching community and hydrological needs through sustainable wetland use. Natural Resources Forum 27, 117–129.
DNRE, 1999. Victorian water quality monitoring network and state biological monitoring programme, manual of procedures. Report prepared for Department of Natural Resources and Environment (WES Report No. 182/97).
Dodson, S.I., 2001. Zooplankton communities of restored depressional wetlands in Wisconsin USA. Wetlands 21, 292–300.
Escalera-Vazquez, L.H., Zambrano, L., 2010. The effect of seasonal variation in abiotic factors on fish community structure in temporary and permanent pools in a tropical wetland. Freshwater Biology 55, 2557–2569.
Feio, M.J., Almeida, S.F.P., Craveiro, S.C., Calado, A.J., 2007. Diatoms and macroinvertebrates provide consistent and complimentary information on environmental quality. Hydrobiologia 169, 247–258.
Feld, C.K., Sousa, J.P., da Silva, P.M., Dawson, T.P., 2010. Indicators for biodiversity and ecosystem services: towards an improved framework for ecosystems assessment. Biodiversity and Conservation 19, 2895–2919.
Fielding, A.H., Bell, J.F., 1997. A review of methods for assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24, 38–49.
Gabriels, W., Goethals, P.L.M., Dedecker, A., Lek, S., De Pauw, N., 2007. Analysis of macrobenthic communities in Flanders, Belgium, using a stepwise input variable selection procedure with artificial neural networks. Aquatic Ecology 41, 427–441.
Goethals, P.L.M., Dedecker, A.P., Gabriels, W., De Pauw, N., 2002. Development and application of predictive river ecosystemmodels based on classification trees and artificial neural networks. In: Recknagel, F. (Ed.), Ecological Informatics: Understanding Ecology by Biologically-Inspired Computation. Springer-Verlag, Berlin, pp. 91–107.
Goethals, P.L.M., Dedecker, A.P., Gabriels, W., Lek, S., De Pauw, N., 2007. Applications of artificial neural networks predicting macroinvertebrates in freshwaters. Aquatic Ecology 41, 491–508.
Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135, 147–186.
Hanson, M.A., Butler, M.G., 1994. Responses to food web manipulation in a shallow waterfowl lake. Hydrobiologia 279 (280), 457–466.
Hey, D.L., Philippi, N.S., 1995. Flood reduction through wetland restoration: the Upper Mississippi River Basin as a case history. Restoration Ecology 3, 4–17.
Hoang, H., Lock, K., Mouton, A., Goethals, P.L.M., 2010. Application of classification trees and support vector machines to model the presence of macroinvertebrates in rivers in Vietnam. Ecology Information 5, 140–146.
Jacobs, A., Rogerson, A., Fillis, D., Bason, C., 2009. Delaware Department of Natural Resources and Environmental Control, Watershed Assessment Section, Dover, Delaware, USA: Wetland condition of the Inland Bays watershed, Volume 1.
Jurado, G.B., Callanan, M., Gioria, M., Baars, J.R., Harrington, R., Kelly-Quinn, M., 2009. Comparison of macroinvertebrate community structure and driving environmental factors in natural and wastewater treatment ponds. Hydrobiologia 634, 153–165.
Kallimanis, A.S., Ragia, V., Sgardelis, S.P., Pantis, J.D., 2007. Using regression trees to predict alpha diversity based upon geographical and habitat characteristics. Biodiversity and Conservation 16, 3863–3876.
Karr, J.R.E., Rossano, E.M., 2001. Applying public health lessons to protect river health. Ecology and Civil Engineering 4, 3–18.
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.
Liston, S.E., Newman, S., Trexler, J.C., 2008. Macroinvertebrate community response to eutrophication in an oligothrophic wetland. An in situ mesocosm experiment. Wetlands 28, 686–694.
Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T., Smith, K.G., Williams, P.H., 2003. Avoiding pitfalls of using species distribution models in conservation planning. Conservation Biology 17, 1591–1600.
Manel, S., Williams, H.C., Ormerod, S.J., 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. Journal of Applied Ecology 38, 921–931.
Mitsch, W.J., Gosselink, J.G., 1993. Wetlands, 2nd edn. Van Nostrand Reinhold, New York.
Mouton, A.M., De Baets, B., Goethals, P.L.M., 2009. Knowledge-based versus data-driven fuzzy habitat suitability models for river management. Environmental Modelling & Software 24, 982–993.
Park, Y.S., Verdonschot, P.F.M., Chon, T.S., Lek, S., 2003. Patterning and predicting aquatic macroinvertebrate diversities using artificial neural network. Water Research 37, 1749–1758.
Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Francisco.
Ramsar, 2010. Wetland ecosystem services. http://www.ramsar.org/2010.
Robertson, M.P., Caithness, N., Villet, M.H., 2001. A PCA-based modelling technique for predicting environmental suitability for organisms from presence records. Division Distributed 7, 15–27.
Rolon, A.S., Maltchik, L., 2006. Environmental factors as predictors of aquatic macrophyte richness and composition in wetlands of southern Brazil. Hydrobiologia 556, 221–231.
Roy, A.H., Rosemond, A.D., Paul, M.J., Leigh, D.S., Wallace, J.B., 2003. Stream macroinvertebrates response to catchment urbanization (Georgia, USA). Freshwater Biology 48, 329–346.
Ságová Marecková, M., Kvet, J., 2002. Impact of oxygen released by the roots of aquatic macrophytes on composition and distribution of benthic macroinvertebrates in a mesocosm experiment. Archives of Hydrobiology 155, 567–584.
Shewaye, D., 2008. Wetlands and management aspects in Ethiopia: situation analysis. Proceedings of the National Stakeholders' Workshop on Creating National Commitment for Wetland Policy and Strategy Development in Ethiopia.
Shine, C., Klemm, C., 1999. Wetlands, Water and the Law: Using Law to Advance Wetland Conservation and Wise Use. IUCN, Gland.
Simonit, S., Perrings, C., 2011. Sustainability and the value of the 'regulating' services: wetlands and water quality in Lake Victoria. Ecological Economics 70, 1189–1199.
Solomon, M., 2004. Socio-Economic Determinants of Wetland Cultivation in Kemise, Illubabor Zone, Southwestern Ethiopia. Eastern Africa Social Science Research Review, 20. Michigan State University Press, pp. 93–114. doi:10.1353/eas.2004.0004.
Steinman, A.D., Conklin, J., Bohlen, P.J., Uzarski, D.J., 2003. Influence of cattle grazing and pasture land use on macroinvertebrate communities in freshwater wetlands. Wetlands 23, 877–889.
Store, R., Kangas, J., 2001. Integrating spatial multi-criteria evaluation and expert knowledge for GIS-based habitat suitability modelling. Landscape and Urban Planning 55, 79–93.
Strand, M., Merritt, R.W., 1999. Impacts of livestock grazing activities on stream insect communities and the riverine environment. American Entomological 45, 13–29.
Teferi, E., Uhlenbrook, S., Bewket, W., Wenninger, J., Simane, B., 2010. The use of remote sensing to quantify wetland loss in the Choke Mountain range, Upper Blue Nile basin, Ethiopia. Hydrology and Earth System Sciences 14, 2415–2428.
ter Braak, C.J.F., Jaap, W., 1994. On the statistical analysis of vegetation change: a wetland affected by water extraction and soil acidification. Journal of Vegetation Science 5, 361–372.
ter Braak, C.J.F., Smilauer, P., 2002. CANOCO reference manual and CanocoDraw for windows user's guide: software for canonical community ordination (version 4.5). , p. 500 (Ithaca, NY, USA).
Trigal-Domínguez, C., Frenández-Aláez, C., García-Criado, F., 2009. Ecological assessment of highly heterogeneous systems: the importance of taxonomic sufficiency. Liminologica 4, 208–214.
Turner, R.K., Morse-Jones, S., Fisher, B., 2010. Ecosystem valuation: a sequential decision support system and quality assessment issues. Annals of the New York Academy of Sciences 1185, 79–101.

Wellborn, G.A., Skelly, D.K., Werner, E.E., 1996. Mechanisms creating community structure across a freshwater habitat gradient. Annual Review of Ecology and Systematics 27, 337–363.

Witten, I.H., Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann Publishers, San Francisco. (369 pp).

Xu, C., Sheng, S., Zhou, W., Cui, L., 2011. Characterizing wetland change at landscape scale in Jiangsu Province, China. Environmental Monitoring and Assessment. doi:10.1007/s10661-010-1735-6.

Yimer, H.D., Mengistou, S., 2009. Water quality parameters and macroinvertebrates index of biotic integrity of the Jimma wetlands, southwest Ethiopia. Journal of Wetlands Ecology 3, 77–93.