Wetlands
https://doi.org/10.1007/s13157-019-01231-1

WETLANDS IN THE DEVELOPING WORLD

Habitat Suitability Modelling of Benthic Macroinvertebrate Community in Wetlands of Lake Tana Watershed, Northwest Ethiopia

Ayenew Gezie[1,2] · Worku Legesse Mulat[3] · Wassie Anteneh[1] · Eshete Dejen[4] · Helmut Kloos[5] · Seid Tiku Mereta[2]

Received: 12 December 2018 / Accepted: 25 September 2019
© Society of Wetland Scientists 2019

Abstract
Predictive modelling corroborates decision-making in the development of a standard habitat assessment protocol. In this study, we modelled environmental requirements of benthic macroinvertebrates. Classification and regression tree models (CART) and ordination analysis were performed to identify important variables affecting macroinvertebrate community pattern in the Lake Tana Watershed. A dataset of 95 samples was collected from eight wetlands. Among the modelled taxa, Coenagrionidae and Libellulidae had substantial predictive performance based on Kappa statistic ($\kappa > 0.6$) whereas Baetidae, Physidae, Tipulidae and Hydrophilidae had moderate predictive model performance ($\kappa \geq 0.4$). Vegetation cover, leather tanning, vegetation clearance and nitrate ion were the topmost selected environmental variables influencing the occurrence of macroinvertebrate taxa. The conditional analysis depicted that the abundance of Coenagrionidae and Libellulidae increased with the increasing in vegetation cover. Overall, macroinvertebrate taxa have a clear habitat requirement within the habitat gradient studied and hence, could be a potential candidate for biomonitoring and provide valuable information in the development of a standard wetland assessment protocol.

Keywords Classification tree · Lake Tana · Macroinvertebrate · Regression tree · Water quality · Wetlands

Introduction

Anthropogenic activities are seriously affecting the ecological integrity of freshwater ecosystems at scales far exceed impacts on natural phenomena (Habersack et al. 2014). Wetlands are freshwater ecosystem most exposed to detrimental human activities despite their contributions to environmental stability. Wetlands play an important role in the sustainability of natural system and human welfare. They perform carbon sequestration, waste treatment, nutrient cycling, flood reduction and are also rich in biodiversity (Costanza et al. 1997; Jacobs et al. 2009; Ramsar 2010; Moomaw et al. 2018). Furthermore, wetlands contribute a crucial role in ensuring safe water supplies, food security and livelihoods for millions of people living in developing countries, including Ethiopia (Teferi et al. 2010; Mereta et al. 2012). Despite their immense contributions, wetlands have been inadequately considered in Ethiopian environmental policy and the country has not ratified Ramsar convention. This is attributed to lack of awareness and logistic constraints (Dixion and Wood 2003; Mereta et al. 2012). Information about the interaction between environmental factors and aquatic fauna is a key issue in conservation management and restoration of wetlands. Therefore, generating sound scientific information and decision support tools may make valuable contributions in aquatic ecosystem management and conservation and analyzing aquatic fauna interactions as a functional environmental variable could be

✉ Seid Tiku Mereta
seidtiku@yahoo.com

1 Department of Biology, Bahir Dar University, P.O. Box 79, Bahir Dar, Ethiopia

2 Department of Environmental Health Sciences and Technology, Jimma University, P.O. Box 378, Jimma, Ethiopia

3 Department of Civil and Environmental Engineering, University of Connecticut, Storrs, CT, USA

4 Intergovernmental Authority on Development, P.O. Box. 2653, Djibouti, Republic of Djibouti

5 Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA

Published online: 19 November 2019

a useful decision support tool (Mereta et al. 2013; Chawaka et al. 2018).

Although considerable attention has been given to ecological assessment, wetland quality deterioration has increased in recent years. It has been estimated that during the last few centuries, about half of all natural wetland ecosystems have been destroyed worldwide due to anthropogenic activities (Xu et al. 2011; Moomaw et al. 2018). Major changes in land use and vegetation/cover at watershed scales are exacerbating the degradation of water quality or loss. In developing countries in particular, this problem needs greater attentions as the nexus between human activities and natural resources is very strong (Teferi et al. 2010; Getachew et al. 2012; Gezie et al. 2017). In Ethiopia, rapid population growth and economic transformations are the main drivers of aquatic resource loss and quality decline (Dixion and Wood 2003; Getachew et al. 2012; Gezie et al. 2017). A number of studies were carried out to understand human impacts on wetland ecosystems and their services in Ethiopia (Dixion 2002; Mereta et al. 2012, 2013; Gezie et al. 2017; Chawaka et al. 2018) and numerous studies pointed out the need for greater attention (Dixion 2002; Dixion and Wood 2003; Gezie et al. 2017). Although decision support tools are useful to provide information for policy and decision makers in developing a standard habitat assessment protocols, they are only gradually being developed in Ethiopia (eg. Beyene et al. 2009; Mereta et al. 2012).

Decision support tools such as classification and regression tree models (CART) provide comprehensive insight for resource management (Guisan and Zimmerman 2000; Guisan and Thuiller 2005; Mereta et al. 2013). CART models are popular and commonly used in ecological studies to assess, monitor and manage ecosystem conditions (Hoang et al. 2010; Chen et al. 2017; Yigezu et al. 2018). CART has a number of advantages over other traditional statistical models. First, it is well suited for analysis of complex ecological data with high–order interactions and captures nonlinear relationship between explanatory and response variables (Breiman et al. 1984; Bilton et al. 2017). Second, it does not rely on the assumptions that are required for parametric statistics and the analysis is not restricted by multicollinearity in predictor variables (Lewis 2013).

It has been reported that various environmental conditions such as vegetation cover, ammonium nitrogen, water pH, hardness, turbidity, nutrients, dissolved oxygen concentration, conductivity and water temperature or various human activities affect the occurrence and abundance of aquatic fauna (Mereta et al. 2012; Yigezu et al. 2018; Yi et al. 2018). However, the relative importance of various environmental variables varies significantly among ecological settings (Yigezu et al. 2018). This implies that local investigations are needed to determine the distribution and habitat requirements of aquatic fauna for habitat conservation purposes. Modelling the distribution of taxa as a function of abiotic environments has been recognized as a significant component of conservation planning (Zhang et al. 2018). Scholars commonly have used macroinvertebrates to develop tools and assess aquatic environment quality as a function of abiotic factors.

Because macroinvertebrate assemblages in freshwater ecosystems integrate the human impacts in the watershed and their assemblages can be considered as indicators of ecosystem status (Habersack et al. 2014). Consequently, knowing the ecological status of freshwater based on the biotic assemblages is an essential prerequisite to aquatic ecosystem assessment, restoration and management (Tsai et al. 2017). Macroinvertebrates are widely studied in the development of decision support tools and biomonitoring because their attributes and ecological roles have been recognized in assessing and monitoring ecosystem impairments (Liston et al. 2008; Feld et al. 2010). Macroinvertebrates link the lower and higher trophic levels in wetland trophic structures (Butkas et al. 2011; Pace et al. 2012). They are also known as an important food source for amphibians, fish and other invertebrates, and they are therefore an integral component of aquatic food webs (Jiang et al. 2010). Therefore, their occurrence, abundance and species richness could indicate level of freshwater degradation (Mereta et al. 2012), and the ability of a wetland ecosystems to support other higher animal taxonomic groups (Batzer et al. 2006). Furthermore, macroinvertebrates are adapted to a wide range of environmental gradients (Gezie et al. 2017; Chawaka et al. 2018). On the other hand, macroinvertebrates are extensively studied, easily visible with the naked eye, and they are taxonomically rich. Because of these characteristics, macroinvertebrate community structures are considered as useful proxies for determining the ecological status of freshwater ecosystems (Li et al. 2012).

Habitat suitability modelling has been recognized as an essential tool to support decision-making in water management (Guisan and Zimmerman 2000). In this context, we aimed to develop decision tree models and ordination analysis to assess the factors determining the occurrence and abundance of macroinvertebrate taxa. The findings obtained in this study used to identify environmental factors that are important for macroinvertebrate community structures in the habitats studied, and used as a guideline in designing management and habitat conservation plans of wetlands and their related ecosystem services in the Lake Tana watershed.

Materials and Methods

Study Area

The study was conducted in Lake Tana watershed. The watershed is located in Amhara National Regional State in northwest Ethiopia. The studied wetlands were Yiganda,

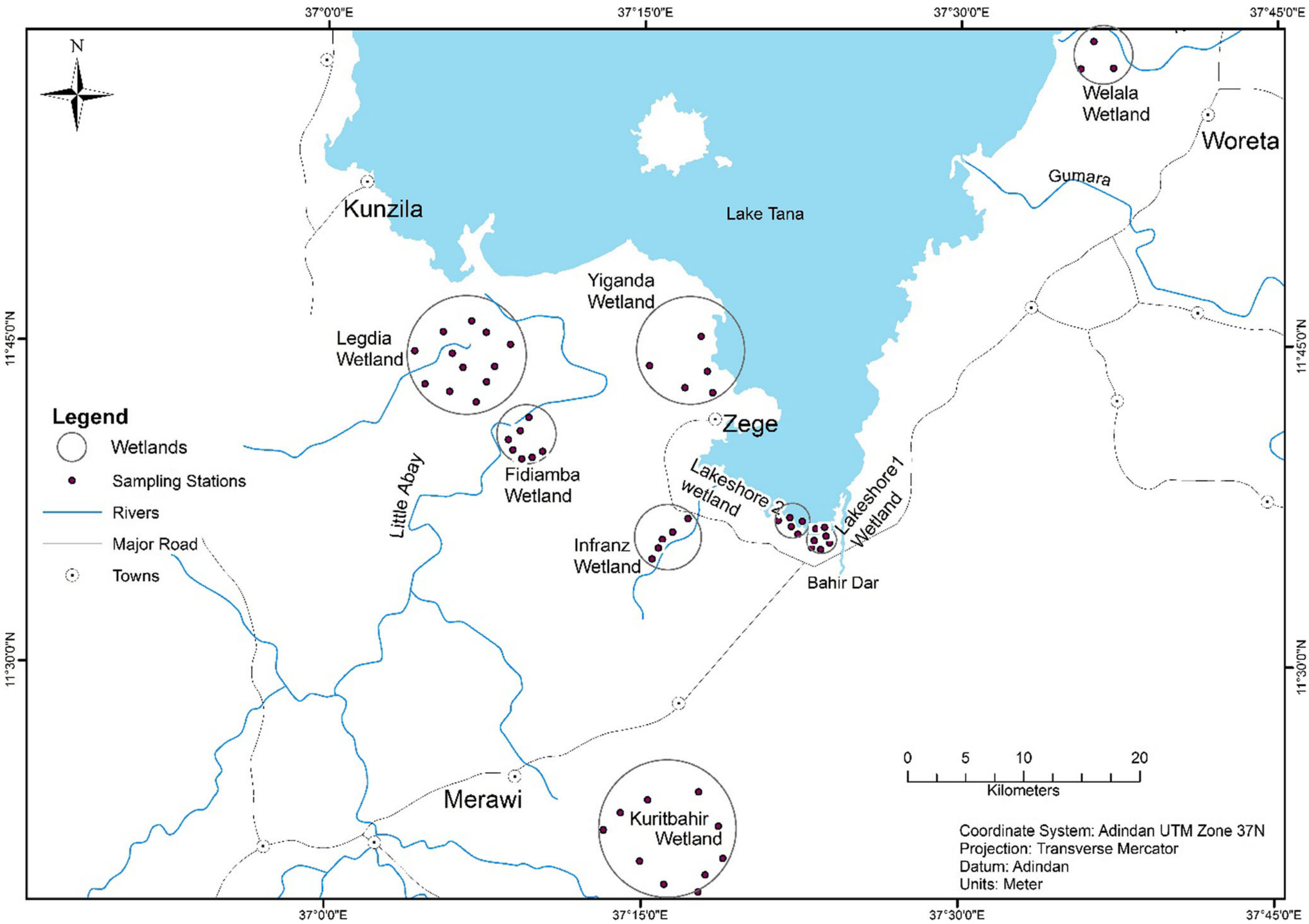

Fig. 1 Location of wetlands and sampling locations in Lake Tana Watershed

Welala, Fidiamba, Infranz, Legdia, Kurtbahir, Lakeshore 1, and Lakeshore 2 (Fig. 1). Data from Welala wetland was collected only in the dry season as it was inaccessible during rainy season. The Yiganda and Welala wetlands connected to Lake Tana during rainy season. Little Abay River flooded Fidiamba wetland in the rainy season, while Infranz wetland is a riparian. Kurtbahir and Legdia wetlands are depressional wetlands (Gezie et al. 2017). Lakeshore 1 and Lakeshore 2 are lacustrine wetlands interfacing Lake Tana with Bahir Dar City, from where they receive both solid and liquid wastes (Wondie 2010; Gezie et al. 2017). The major anthropogenic activities observed in and around the study wetlands were farming, leather tanning and processing, intensive grazing, drainage, vegetation clearance, water abstraction, waste dumping, and eucalyptus plantation (Gezie et al. 2018).

Data Collection

Data were collected at 42 sampling sites in dry season from April to May 2015 from eight wetlands while data were collected in rainy season at 53 sampling sites from September to October 2015 from sites located in seven wetlands. A total of 95 samples were collected both in dry and rainy seasons. Sites were selected within each wetland along a gradient of visible disturbance including both less disturbed and heavily disturbed sites. The number of sampling sites was distributed among the wetlands according to their size, with the smallest wetlands having a lower number of sampling sites. The accessibility of the wetlands was also taken in to consideration. Habitat characteristics at each sampling site were assessed following the USEPA wetland habitat assessment protocol (Baldwin et al. 2005). The level of human perturbations in regard to hydrological modifications, habitat alteration, and land use practices was assessed. Habitat alterations included farming, livestock grazing, tree planting, and vegetation clearance, and, waste dumping and leather tanning were the major polluting activities in the area. We used the protocol described by Hruby (2004), and modified by Mereta et al. (2013) to quantify anthropogenic perturbations. The magnitude of each of the disturbance was quantified on an ordinal scale. A score of 1 was assigned to no or minimal disturbance, 2 to moderate, and 3 to high disturbances. The overall disturbance for each site was calculated by summing the individual values of disturbance factors (eight different factors in total).

Table 1 Input variables used for the model development: mean values, standard deviation, and range: COD: Chemical Oxygen Demand

Variable	Unit	Mean ± SD	Range
Ammonium	mg/l	0.04 ± 0.09	0.001–0.776
Nitrate	mg/l	1.38 ± 1.26	0.001–4.50
Total nitrogen	mg/l	13.83 ± 8.81	0.892–44.4
Orthophosphate	mg/l	0.25 ± 0.27	0.010–1.31
Total phosphorus	mg/l	0.54 ± 0.53	0.060–3.04
Chlorophyll a	μ g/l	14.83 ± 8.06	11.310–63.2
COD	mg/l	88.24 ± 88.55	1.390–37
Conductivity	μS/cm	300.58 ± 247.32	1.980–1641
Dissolved oxygen	mg/l	4.54 ± 2.36	0.750–13.88
Turbidity	NTU	143.66 ± 104.21	10.500–745
pH	–	7.31 ± 0.80	3.210–9.73
Vegetation cover	%	71.37 ± 23.91	0.000–95
Sludge thickness	cm	20.31 ± 21.46	0–1
Secchi depth	cm	20.75 ± 16.91	2.00–110
Water temperature	°C	24.52 ± 3.97	7.25–33.5
Plantation	N/A	1.97 ± 0.93	1–3
Grazing	N/A	1.93 ± 0.91	1–2
Vegetation clearance	N/A	2.32 ± 0.84	1–3
Farming	N/A	2.08 ± 0.87	1–3
Drainage	N/A	2.04 ± 0.89	1–3
Abstraction	N/A	1.82 ± 0.82	1–3
Waste dumping	N/A	1.75 ± 0.87	1–3
Leather tanning	N/A	1.78 ± 0.91	1–3

Abiotic Habitat Characteristics

Water depth and sludge layer thickness were measured at each sampling site using a graduated stick. Water conductivity, pH, dissolved oxygen, 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 used as a proxy of phytoplankton biomass and measured in the field using a handheld fluorometer (Turner Design Aqua Fluor). From each site, a water sample (200 ml) was taken and filtered through a 0.45-μm filter paper in the field for the determination of nitrate, ammonia, and orthophosphate concentration. Unfiltered water samples (500 ml) were collected from each site to determine total organic nitrogen (TON), total phosphorous (TP) and the chemical oxygen demand (COD) concentrations in the laboratory. Water samples were kept cool in the dark during transportation to Jimma University environmental health laboratory. Ammonia was analyzed using direct nesslerization method (APHA 1998). Total phosphorus samples were first digested in a block digester using ammonium persulfate and sulfuric acid reagent (APHA 1998). Samples for TON and COD were digested and measured with photometric kits (HACH LANGE) using a Hach DR5000 spectrophotometer. The percentage of vegetation cover was visually estimated within a 500-m radius around each observation site (Baldwin et al. 2005).

Biotic Habitat Characteristics

Ninety-five macroinvertebrate samples were collected from a total of 53 sampling sites. A rectangular frame net (20 × 30 cm) with a mesh size of 300 μm was used to kick for 5 min along a 10-m stretch per site (Gabriels et al. 2010). The bottom sediment was disturbed by foot to increase the probability of catching macroinvertebrates through the kick net. The collected invertebrates were sorted in the field and stored in 80% ethanol. Then, the invertebrate samples were transported to the laboratory and examined using a stereomicroscope (×10 magnification). Identification of invertebrates was performed to the family level using the identification key developed by Bouchard (2012).

Data Analysis

Multivariate statistical analysis and decision tree models such as classification and regression tree models were used to analyze the habitat preference of macroinvertebrate taxa in the studied wetlands.

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

Family	Frequency of occurrence (%)	Relative abundance
Gomphidae	21	226
Mesoveliidae	23	102
Sphaeriidae	28	108
Culicidae	29	90
Tipulidae	29	60
Naucoridae	31	230
Nepidae	34	104
Physidae	40	214
Coenagrionidae	47	1096
Libellulidae	47	931
Planorbidae	48	296
Notonectidae	49	820
Aeshnidae	53	254
Baetidae	57	578
Corixidae	63	776
Lymnaeidae	68	809
Chironomidae	69	1057
Belostomatidae	72	979
Hydrophilidae	77	1628

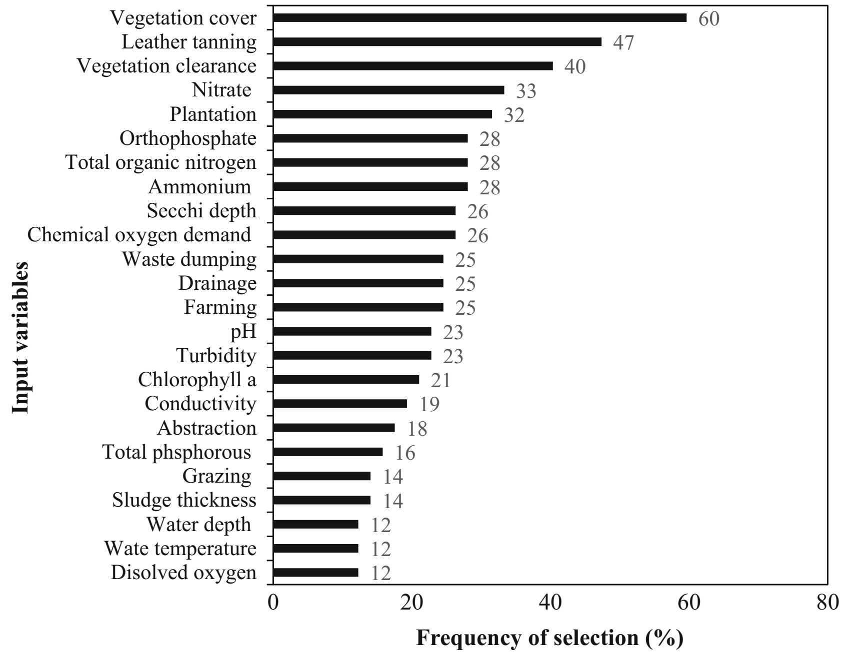

Fig. 2 Overview of the average frequency of selection of the input variables used in constructing the classification tree models

Classification Tree Models

Twenty-five environmental variables (Table 1) and 19 most frequently occurring macroinvertebrate taxa (Table 2) were used to investigate species-environment relationships in the study area. Families occurred in more than 20% of the samples were included in the CART model. Both classification and regression tree models were applied to induce the decision tree models. J48 algorithm (Quinlan 1993) was applied to develop the classification tree models in WEKA (Witten and Frank 2005). Regression trees models were developed using the M5 algorithm in WEKA (Witten and Frank 2005) to relate the abundance of macroinvertebrate taxa to selected environmental variables.

Training and validation of the classification tree models were based on a three-fold cross validation procedure (Witten and Frank 2005). Both dry and rainy season data were used to construct the models. Prior to the analysis, all data were randomized and the dataset was stratified into three subsets. Thereafter, two-thirds of the dataset were used as training data while one-third of the dataset was used for testing the model. The cross-validation processes were repeated three times using one of the three subsets as a validation dataset once. In this way, three models were built for each macroinvertebrate taxa. 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 outcomes of the three models.

The predictive performance of the models was evaluated based on Cohen's Kappa statistic (k)(Cohen 1960) and the percentage of correctly classified instances (CCI)(Witten and Frank 2005). Classification tree models with a value CCI equal to or higher than 70% and kappa value higher than or equal to 0.4 were considered reliable (Dakou et al. 2007). We used the ranges of Kappa recommended by Landis and Koch (1977) for model performance evaluation: $\kappa \leq 0$ (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).

Conditional Analysis

Conditional analysis was performed to gain insight into the response of macroinvertebrate taxa occurrence to input environmental variables. For each of the macroinvertebrate taxa, three regression tree models were constructed per taxon. This was done by plotting each 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 (Mereta et al. 2012; Yigezu et al. 2018). In this way, a line was plotted showing the relationships between the environmental factors and the abundance of macroinvertebrates. The determination coefficient (R^2) value was used 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 models (Kallimanis et al. 2007).The closer the value to one, the better the model performed.

Multivariate Data Analysis

Canonical correspondence analysis (CCA) was performed using CANOCO 4.5 (ter Braak and Smilauer 2002) to

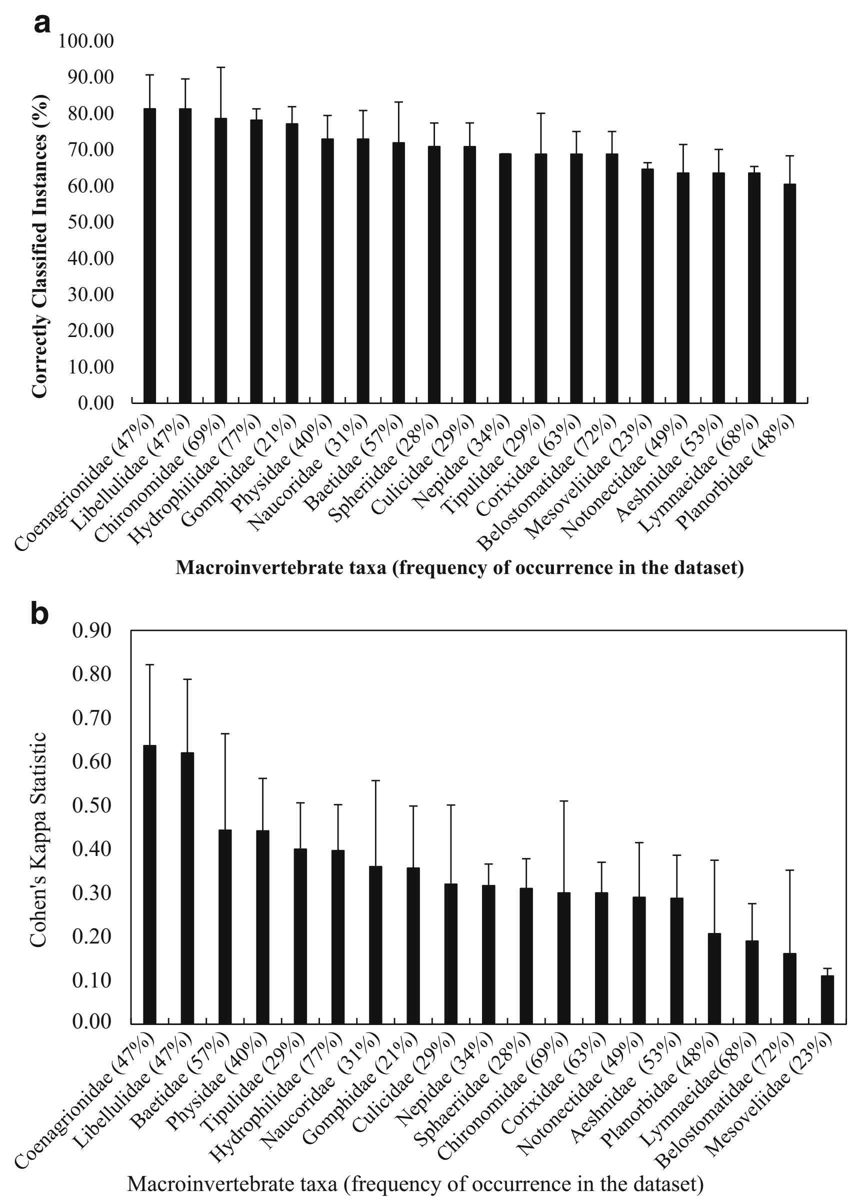

Fig. 3 Overview of the average predictive performance of each taxa based on **a** Correctly Classified Instances and **b** Cohen's Kappa statistic. The frequency of occurrence of each taxon is given in each column

investigate the species-environment relationships on ordination axes. Nineteen invertebrate taxa were included in this ordination analysis. The environmental parameters were log transformed log(x + 1) in the analysis to obtain homogeneity of variance in Canonical correspondence analysis. The statistical significance of eigenvalues and species–environment correlations generated by the CCA were tested using Monte Carlo permutations. All data (dry and rainy season sampling data) were used together to construct the plots. Prior to the ordination analysis, a pre-selection was carried out to remove those variables with relatively high multi-collinearity. Multi-collinearity was assessed by examining variance inflation factor (VIF)(ter Braak and Smilauer 2002). When two or more variables had VIF of larger than 5, one of the variables was included in the model.

Results

Variable of Importance

Table 1 shows the predictor variables for determining the presence/absence of 19 benthic macroinvertebrate taxa while Table 2 shows frequency of occurrence of macroinvertebrates in all samples and their total abundance. Figure 2 depicted the average frequency of selection of the predicator variables by the classification tree models on the different models. Since the training and validation were based on three-fold cross validation, three models were developed for each taxon.

The most frequently selected input variable by the classification tree models was vegetation cover (60%) followed leather tanning (47%), vegetation clearance (40%) and nitrate

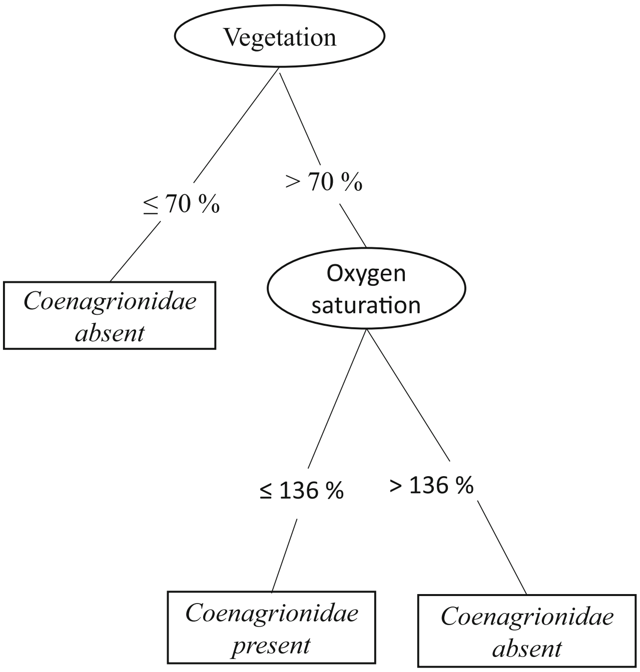

Fig. 4 Classification trees model predicting the presence or absence of *Coenagrionidae* (Correctly Classified Instances = 90.63%, Kappa = 0.82)

(33%) (Fig. 2). On the other hand, water depth, water temperature and dissolved oxygen were the least selected input variables by classification tree models (Fig. 2).

Model Performance Evaluation

The model performances based on the average CCI and Cohen's Kappa statistic of the three-fold cross validation for 19 macroinvertebrate taxa are shown in Fig. 3a and b. The error bars indicate the variation between the subset models constructed per taxon. The average CCI varied between 60.43 ± 7.86% and 81.26 ± 9.37%. Based on CCI and Cohen's Kappa statistic, Coenagrionidae, Libellulidae, Hydrophilidae, Gomphidae, Physidae, and Naucoridae had reliable classification tree models (CCI >70%; $\kappa > 0.40$) (Fig. 3a). On the other hand, Coenagrionidae and Libellulidae had substantial predictive model performance with Cohen's Kappa statistic ($\kappa > 0.60$). Whereas, Baetidae, Physidae, Tipulidae and Hydrophilidae had moderate predictive model performance with Cohen's Kappa statistic ($\kappa \geq 0.40$). Lymnaeidae, Belostomatidae, and Mesoveliidae slight model performance with Cohen's Kappa statistic of 0.19 ± 0.09, 0.16 ± 0.19 and 0.11 ± 0.02, respectively (Fig. 3b). Whereas the rest of the macroinvertebrates modeled in this study (Fig. 3b) had fair model performance.

Among the classification trees developed for the nineteen macroinvertebrate taxa, only trees with high predictive performance, more transparent and ecologically meaningful were presented. An example of classification tree models for five macroinvertebrates is shown in Figs. 4, 5, 6, 7 and 8. The

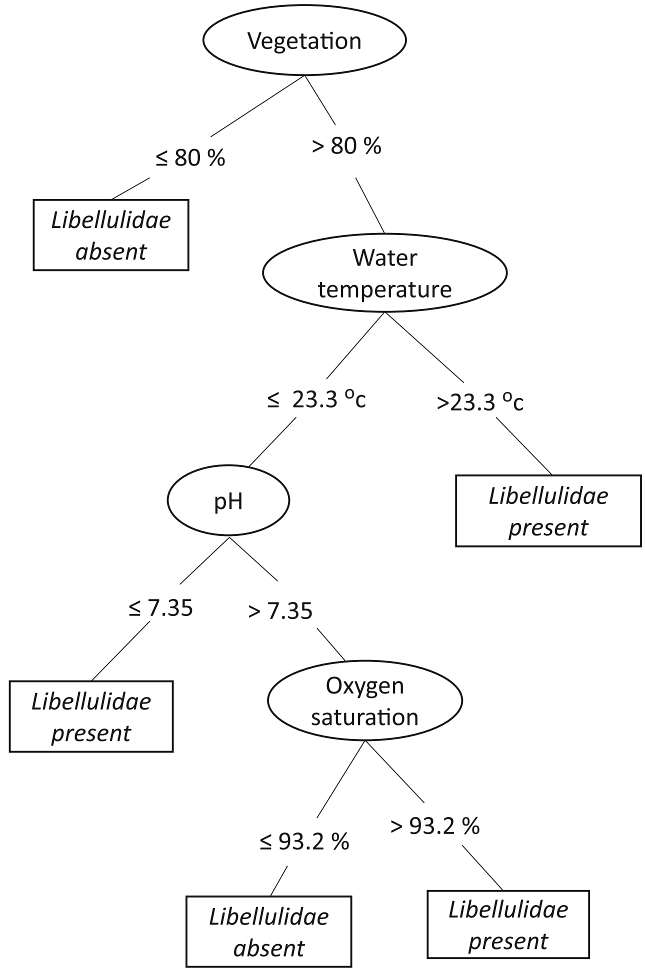

Fig. 5 Classification trees model predicting the presence or absence of *Libellulidae* (Correctly Classified Instances =78.13%, Kappa = 0.56)

classification trees indicate that vegetation cover was given as the root of the trees attribute to predict the occurrence of Coenagrionidae (Fig. 4), Libellulidae (Fig. 5), Nepidae

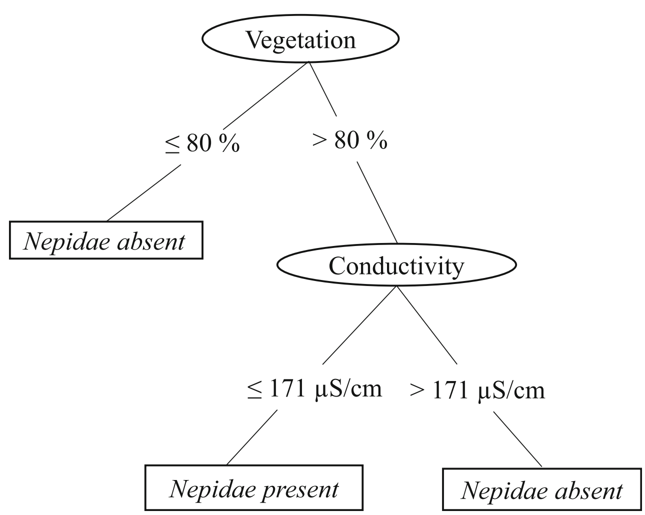

Fig. 6 Classification trees model predicting the presence or absence of *Nepidae* (Correctly Classified Instances = 68.8%, Kappa = 0.34)

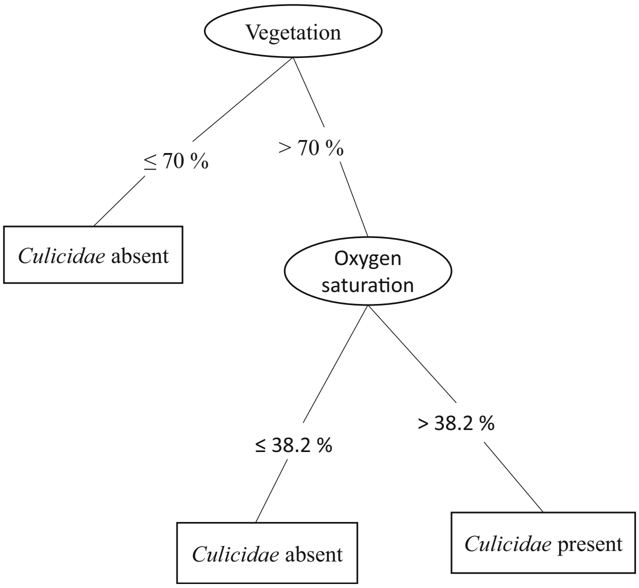

Fig. 7 Classification trees model predicting the presence or absence of *Culicidae* (Correctly Classified Instances = 78.13%, Kappa = 0.51)

(Fig. 6) and Culicidae (Fig. 7). Leather tanning was the most informative attribute given as the root of the trees to predict the occurrence of Baetidae (Fig. 8).

Conditional Analysis

The selection of taxa for conditional analysis was based on its predictive performance, complexity of the tree and the type of variables selected. The conditional analysis based on regression tree models for Coenagrionidae, Libellulidae, and Baetidae is shown on Fig. 9. The determination coefficients obtained from the regression tree model ranged from 0.329 ± 0.12 to 0.517 ± 0.19. Conductivity and vegetation cover were important predictor of the abundance of macroinvertebrate taxa. The regression tree models indicated that the abundance

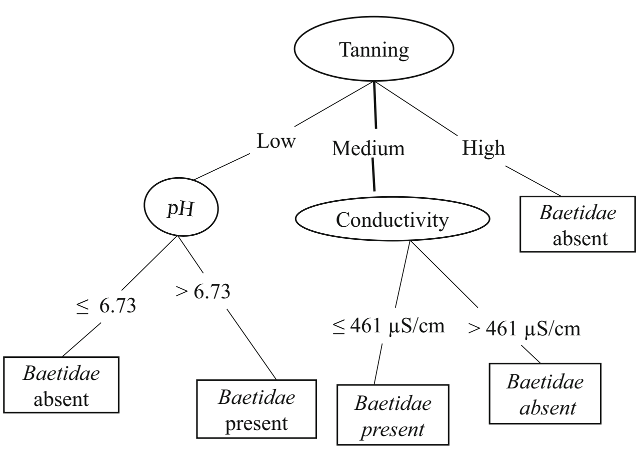

Fig. 8 Classification trees model predicting the presence or absence *Baetidae* (Correctly Classified Instances = 75%, Kappa = 0.5)

of Coenagrionidae and Libellulidae increased with increasing vegetation cover (Fig. 9a, b). The abundance of Baetidae and Coenagrionidae also declined with increasing in water conductivity (Fig. 9c, d).

Ordination Analysis

The first and second canonical axes explained 16.9% and 8.2% of the variances in the species data, respectively. The species–environment correlation of the first axis was statistically significant in a Monte Carlo permutation test ($P < 0.05$). Vegetation cover, total phosphorous, chlorophyll a, leather tanning, pH and Secchi depth significantly explained the variance of species-environment relationship on the ordination axes ($p < 0.05$) (Fig. 10). Whereas, the first axis of CCA was positively correlated with sludge layer thickness and nitrate and it is negatively correlated with vegetation cover. CCA axis 2 was positively correlated with total phosphorous, orthophosphate, chlorophyll a as well as leather tanning, but it negatively correlated with secchi depth and water depth (Fig. 10).

Discussion

In the present study, decision tree models and ordination analysis were used to identify factors influencing the occurrence and abundance of macroinvertebrate taxa in the wetlands of Lake Tana watershed. Vegetation cover, leather tanning, vegetation clearance and nitrate were found to be the topmost environmental factors determining macroinvertebrate taxa occurrence (Fig. 2). Using Kappa values as indicator of the accuracy of classification tree models, Coenagrionidae and Libellulidae had substantial predictive model performances, whereas Baetidae, Physidae, Tipulidae and Hydrophilidae had moderate predictive performance (Fig. 3b). These results indicate that these taxa have distinct habitat quality requirements within the habitat gradient studied. However, the classification tree models performed least for Lymnaeidae, Belostomatidae and Mesoveliidae, suggesting that other factors than the ones we quantified determined their distributions. The lower kappa value and the higher CCI of the model for Belostomatidae may also be ascribed by its high frequency of occurrence while the low kappa value for Mesoveliidae and Tipulidae may be related to their low frequency of occurrence at the study sites, indicating the predictions may be easily generated by chance (Dedecker et al. 2007; Mereta et al. 2012). There is a probability that the most common taxa at the study sites are always present and the rarest taxa are always absent in classification trees models (Dedecker et al. 2007).

Environmental variables clearly affected the occurrence of macroinvertebrate families in the study area. Vegetation cover, leather tanning, vegetation clearance and nitrate were the most

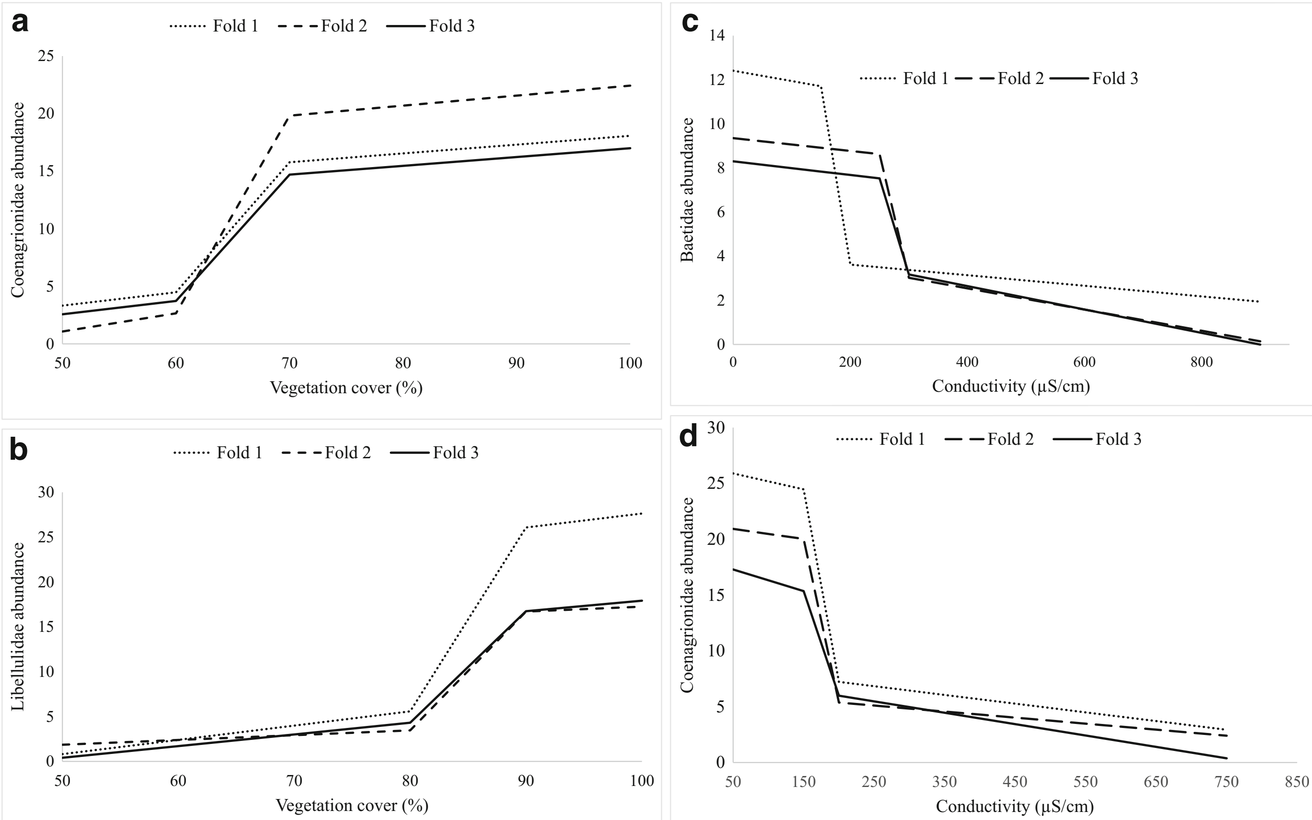

Fig. 9 Conditional analysis illustrating the abundance (number of individuals per sample) of **a** *Coenagrionidae* as a function of vegetation cover (r = 0.517 ± 0.19), **b** *Libellulidae* as a function of vegetation cover (0.430 ± 0.03), **c** *Baetidae* as a function of water conductivity (r = 0.329 ± 0.12), **d** abundance of *Coenagrionidae* as a function of water conductivity (r = 0.465 ± 0.243)

relevant environmental variables influencing the occurrence of macroinvertebrates. Vegetation cover ranked at the roots of the classifications trees evaluating the presence/absence of Coenagrionidae, Libellulidae, Nepidae and Culicidae, implying that this variable is among important requirements for the occurrence of these taxa. Vegetation cover was strongly correlated with the patterns of macroinvertebrate community structure on the ordination axes (Fig. 10). Many macroinvertebrate species depend on vegetation cover for egg laying, nutrition, while predator species need vegetation for successful hunting strategies, thereby directly and positively impacting on macroinvertebrate community compositions (Couceiro et al. 2007; Bloechl et al. 2010).

Besides vegetation cover, water conductivity was an important environmental variable influencing the abundances of macroinvertebrate taxa. Kefford (1998) found water conductivity to be most important in terms of describing the structure of macroinvertebrate communities. In the present study, some tolerant taxa, such as Planorbidae and Lymnaeidae were frequently occurring in water with high electric conductivity. The relative abundance of Chironomidae was also positively related with water conductivity. Chironomidae is a ubiquitous family encompassing a number of species showing broad ecological sensitivity (Panatta et al. 2007). Mereta et al. (2012) reported high abundance of Chironomidae at sites with high water conductivity. In contrast, Nepidae, Naucoridae and Baetidae were encountered in habitats with low concentration of water electric conductivity. Similarly, the conditional analysis results showed that the abundance of Coenagrionidae dramatically declined when water conductivity was higher than 200 µS/cm. Several studies reported that urbanization can contribute to increased levels of conductivity in freshwater ecosystems mainly due to the liberation of ions through decomposition of oxygen demanding wastes (Roy et al. 2003).

Human activities such as leather tanning were also important factor affecting the occurrence of Baetidae. Leather tanning ranked at the roots of the classifications trees built for Baetidae. Leather tanning is widely practiced in wetlands to remove the flesh and fur of animals (Gezie et al. 2017). The process of leather tanning affects water quality as it releases biodegradable organic materials such as proteins and carbohydrates. Microbial decomposition of organic matters results in depletion dissolved oxygen (Mwinyihij et al. 2006). Thus, leather tanning can degrade habitat conditions and water quality, thereby reducing biodiversity. The results also showed that vegetation clearance and nitrate ion concentration

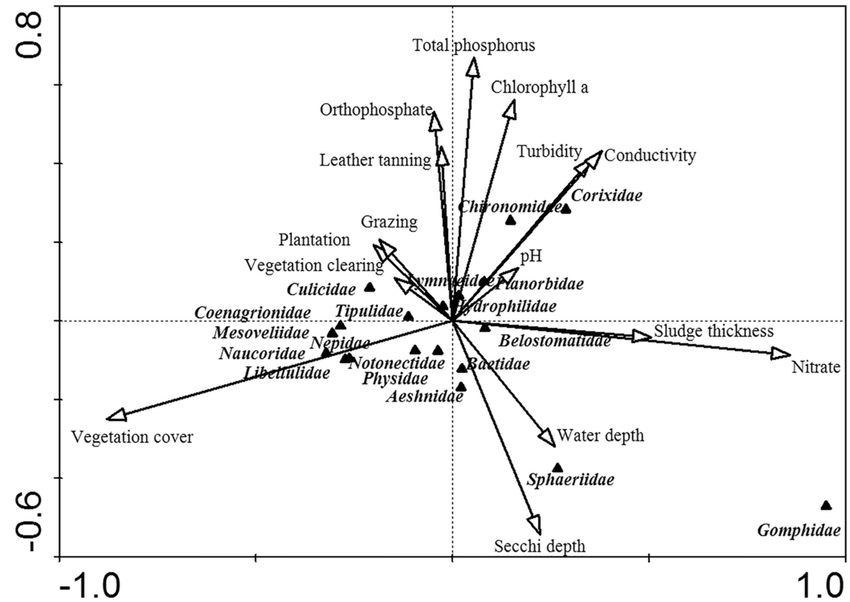

Fig. 10 Canonical correspondence analysis (CCA) of macroinvertebrate taxa and environmental variables in wetlands of northwest Ethiopia

affected the presence or absence of macroinvertebrate families in the studied wetlands. Vegetation clearance alters physical habitat conditions whereas nutrient enrichment causes eutrophication, which in turn deteriorate water quality and influences macroinvertebrate compositions (Kasangaki et al. 2008). Concentrations of nutrients are among the most common causes of declining water quality and physical habitat deterioration (Moal et al. 2018). These impacts alter both water chemistry and macroinvertebrate community compositions (Ambelu et al. 2010).

In conclusion, macroinvertebrates revealed clear interactions to a wide range of environmental variable. Hence, macroinvertebrate taxa could be potential candidates for biomonitoring and could provide valuable insight in the development of a standard wetland assessment protocol. Decision tree models and ordination analysis identified environmental variables influencing the structure of macroinvertebrate communities. Habitat alterations, habitat modifications, land use practices and water quality were important factors in determining macroinvertebrate community patterning in the Lake Tana Watershed. Habitat modifications such as vegetation clearance and improper management of solid and liquids wastes contributes to the deterioration of water quality by raising water electric conductivity and nutrient enrichment levels and thus, decline in aquatic biodiversity (Moal et al. 2018). Therefore, a minimal preservation of wetland vegetation and proper management of solid and liquid wastes are essential to maintain a high biodiversity and to protect ecosystem services in wetlands of Lake Tana watershed.

Authors' Contribution AG conceived the main idea of the paper, collected the data, and prepared the draft. WA, ES, WLM and HK helped in writing the paper. STM assisted data analysis and writing the paper.

Funding Information This research was supported by Jimma University and Bahir Dar University.

Compliance with Ethical Standards

Competing Interests The authors declare that they have no competing interests.

References

Ambelu A, Lock K, Goethals PLM (2010) Comparison of modelling techniques to predict macroinvertebrate community composition in rivers of Ethiopia. Ecological Informatics 5:147–152

APHA (1998) Standard methods for the analysis of wastewater, 20th edn. American Public Health Association, Washington, DC

Baldwin DS, Nielsen DL, Bowen PM, Williams J (2005) Recommended methods for monitoring floodplains and wetlands. 1921038 20 9. (MDBC Publication No. 72/ 04)

Batzer DP, Cooper R, Wissinger SA (2006) Wetland animal ecology. In: Batzer DP, Sharitz RR (eds) Ecology of freshwater and estuarine wetlands. University of California Press, Berkeley, pp 242–284

Beyene A, Addis T, Kifle D, Legesse W, Kloos H, Triest L (2009) Comparative study of diatoms and macroinvertebrates as indicators of severe water pollution: case study of the Kebena and Akaki rivers in Addis Ababa, Ethiopia. Ecological Indicators 9:381–392

Bilton P, Jones G, Ganesh S, Haslett S (2017) Classification trees for poverty mapping. Computational Statistics and Data Analysis 115: 53–66

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

Bouchard RW (2012) Guide to aquatic invertebrate families of Mongolia. Identification manual for students, citizen monitors, and aquatic resource professionals. Jr. Saint Paul, Minnesota

Breiman L, Friedman JH, Olshen RA, Stone CG (1984) Classification and regression trees. Wadsworth International Group, Belmont

Butkas KJ, Vadeboncoeur Y, Vander Zanden MJ (2011) Estimating benthic invertebrate production in lakes: a comparison of methods and scaling from individual taxa to the whole-lake level. Aquatic Sciences 73:153–169

Chawaka SN, Boets P, Mereta ST, Ho LT, Goethals PLM (2018) Using macroinvertebrates and birds to assess the environmental status of wetlands across different climatic zones in southwestern Ethiopia. Wetlands 38(4):653–665

Chen W, Xie X, Wang J, Pradhan B, Hong H, Bui DT, Duan Z, Ma J (2017) A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. Catena 151:147–160

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 RV, Paruelo J, Raskin RG, Sutton P, van den Belt M (1997) The value of the world's ecosystem services and natural capital. Nature 387:253–259

Couceiro SRM, Hamada N, Luz SLB, Forsberg RB, Pimentel TP (2007) Deforestation and sewage effects on aquatic macroinvertebrates in urban streams in Manaus, Amazonas, Brazil. Hydrobiologia 575: 271–284

Dakou E, D'heygere T, Dedecker AP, Goethals PLM, Dimitriadou ML, 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 KE (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81:3178–3192

Dedecker A, Van Melckebeke K, Goethals PLM, 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

Dixion AB (2002) The hydrological impacts and sustainability of wetlands drainage cultivation in Illubabour Ethiopia. Land Degradation and Development 13:17–31

Dixion AB, Wood AP (2003) Wetland cultivation and hydrological management in eastern Africa: matching community and hydrological needs through sustainable wetland use. Natural Resources Forum 27:117–129

Feld CK, Sousa JP, da Silva PM, Dawson TP (2010) Indicators for biodiversity and ecosystem services: towards an improved framework for ecosystems assessment. Biodiversity and Conservation 19: 2895–2919

Gabriels W, Lock K, De Pauw N, Goethals PLM (2010) Multimetric macroinvertebrate index Flanders (MMIF) for biological assessment of rivers and lakes in Flanders (Belgium). Limnologica 40:199–207

Getachew M, Ambelu A, Mereta ST, Legesse W, Adugna A, Kloose H (2012) Ecological assessment of Cheffa wetland in the Borkena Valley, Northeast Ethiopia: macroinvertebrate and bird communities. Ecological Indicators 15:63–71

Gezie A, Anteneh W, Dejen E, Mereta ST (2017) Effects of human-induced environmental changes on benthic macroinvertebrate assemblages of wetlands in Lake Tana watershed, Northwest Ethiopia. Environmental Monitoring and Assessment 189:152

Gezie A, Assefa WW, Getnet B, Anteneh W, Dejen E, Mereta ST (2018) Potential impacts of water hyacinth invasion and management on water quality and human health in Lake Tana watershed, Northwest Ethiopia. Biological Invasion 20(9):2517–2534

Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. Ecology Letters 8:993–1009

Guisan A, Zimmerman NE (2000) Predictive habitat distribution models in ecology. Ecological Modeling 135:147–186

Habersack H, Haspe D, Kondolf M (2014) Large rivers in the Anthropocene – insights and tools for understanding climatic, land use, and reservoir influences. Water Resources Research 50:3641–3646

Hoang TH, Lock K, Mouton A, Goethals PLM (2010) Application of classification trees and support vector machines to model the presence of macroinvertebrates in rivers in Vietnam. Ecological Informatics 5:140–146

Hruby T (2004) Washington State wetland rating system for eastern Washington. Washington State Department of Ecology, Publication No 04-06-15

Jacobs A, Rogerson A, Fillis D, Bason C (2009) Delaware department of natural resources and environmental control. Watershed Assessment Section, Dover

Jiang XM, Xiong J, Qiu JW, Wu JM, Wang JW, Xie ZC (2010) Structure of macroinvertebrate communities in relation to environmental variables in a subtropical Asian River system. International Review of Hydrobiology 95:42–57

Kallimanis AS, Ragia V, Sgardelis SP, Pantis JD (2007) Using regression trees to predict alpha diversity based upon geographical and habitat characteristics. Biodiversity and Conservation 16:3863–3876

Kasangaki A, Chapman LJ, Balirwa J (2008) Land use and the ecology of benthic macroinvertebrate assemblages of high-altitude rainforest streams in Uganda. Freshwater Biology 53:68–697

Kefford BJ (1998) The relationship between electrical conductivity and selected macroinvertebrate communities in four river systems of south-West Victoria, Australia. International Journal of Salt Lake Research 7:153–170

Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174

Lewis RJ (2013) An introduction to classification and regression tree (CART) analysis. California: Presented at the 2000 Annual Meeting of the Society for Academic Emergency Medicine in San Francisco. https://pdfs.semanticscholar.org/6d4a/347b99d056b7b1f28218728f1b73e64cbbac.pdf Accessed 25 November, 2018

Li F, Chung N, Bae MJ, Kwon YS, Park YS (2012) Relationships between stream macroinvertebrates and environmental variables at multiple spatial scales. Freshwater Biology 57:2107–2124

Liston SE, Newman S, Trexler JC (2008) Macroinvertebrate community response to eutrophication in an oligothrophic wetland. An in situ mesocosm experiment. Wetlands 28:686–694

Mereta ST, Boets P, Bayih AA, Malu A, Ephrem Z, Sisay A, Endale H, Yitbarek M, Jemal A, De Meester L, Goethals PLM (2012) Analysis of environmental factors determining the abundance and diversity of macroinvertebrate taxa in natural wetlands of Southwest Ethiopia. Ecological Informatics 7:52–61

Mereta ST, Boets P, De Meester L, Goethals PLM (2013) Development of a multimetric index based on benthic macroinvertebrates for the assessment of natural wetlands in Southwest Ethiopia. Ecological Indicators 29:510–521

Moal ML, Gascuel-Odoux C, Ménesguen A, Souchon Y, Étrillard C, Levain A, Moatar F, Pannard A, Souchu P, Lefebvre A, Pinay G (2018) Eutrophication: a new wine in an old bottle? Science of the Total Environment 651:1–11

Moomaw WR, Chmura GL, Davies GT, Finlayson CM, Middleton BA, Natali SM, Perry JE, Roulet N, Sutton-Grier AE (2018) Wetlands in a changing climate: science, policy and management. Wetlands 38(2):183–205

Mwinyihij M, Strachan NJC, Dawson J, Meharg A, Killham K (2006) Anecotoxicological approach to assessing the impact of tanning industry effluent on river health. Archives of Environmental Contamination and Toxicology 50:316–324

Pace G, Bella VD, Barile M, Andreanic P, Mancini L, Belfiore C (2012) A comparison of macroinvertebrate and diatom responses to anthropogenic stress in small sized volcanic siliceous streams of Central Italy (Mediterranean ecoregion). Ecological Indicators 23:544–554

Panatta A, Stenert C, Santos EMD, Maltchik L (2007) Diversity and distribution of Chironomid larvae in wetlands in southern Brazil. Journal of the Kansas Entomological Society 80:229–242

Quinlan JR (1993) C4.5: programs for machine learning. Morgan Kaufmann Publishers, San Francisco

Ramsar (2010) Wetland ecosystem services. http://www.ramsar.org/2010. Accessed 20 Nov 2018

Roy AH, Rosemond AD, Paul MJ, Leigh DS, Wallace JB (2003) Stream macroinvertebrate response to catchment urbanization (Georgia, U.S.a.). Freshwater Biology 48:329–346

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 CJF, Smilauer P (2002) CANOCO reference manual and CanocoDraw for windows user's guide: software for canonical community ordination (version4.5). PP. 500 (Ithaca, NY, USA)

Tsai WP, Huang SP, Cheng ST, Shao KT, Chang FJ (2017) A data-mining framework for exploring the multi-relation between fish species and water quality through self-organizing map. Science of the Total Environment 579:474–483

Witten IH, Frank E (2005) Data mining: practical machine learning tools and techniques with Java implementations. Morgan Kaufmann Publishers, San Francisco (369 pp)

Wondie A (2010) Improving management of shoreline and riparian wetland ecosystems: the case of Lake Tana catchment. Ecohydrology and Hydrobiology 10:123–132

Xu C, Sheng S, Zhou W, Cui L (2011) Characterizing wetland change at landscape scale in Jiangsu Province, China. Environmental Monitoring and Assessment 179(1-4):279–292

Yi Y, Sun J, Yang Y, Zhou Y, Tang C, Wang X, Yang Z (2018) Habitat suitability evaluation of a benthic macroinvertebrate community in a shallow lake. Ecological Indicators 90:451–459

Yigezu G, Mandefro B, Mengesha Y, Yewhalaw D, Beyene A, Ahmednur M, Abdie Y, Kloos H, Mereta ST (2018) Habitat suitability modelling for predicting potential habitats of freshwater snail intermediate hosts in Omo-gibe river basin, Southwest Ethiopia. Ecological Informatics 45:70–80

Zhang M, Muñoz-Mas R, Martinez-Capel F, Qu X, Zhang H, Peng W, Liu X (2018) Determining the macroinvertebrate community indicators and relevant environmental predictors of the Hun-Tai River basin (Northeast China): a study based on community patterning. Science of the Total Environment 634:749–759

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.