Document downloaded from:

http://hdl.handle.net/10251/128741

This paper must be cited as:

Muñoz Mas, R.; Sánchez-Hernández, J.; Mcclain, M.; Tamatamah, R.; Mukama, SC.; Martinez-Capel, F. (2019). Investigating the influence of habitat structure and hydraulics on tropical macroinvertebrate communities. Ecohydrology & Hydrobiology. 19(3):339-350. https://doi.org/10.1016/j.ecohyd.2018.07.005

The final publication is available at

http://doi.org/10.1016/j.ecohyd.2018.07.005

Copyright Elsevier

Additional Information

INVESTIGATING THE INFLUENCE OF HABITAT STRUCTURE AND HYDRAULICS ON TROPICAL MACROINVERTEBRATE COMMUNITIES

Rafael Muñoz-Mas[1], Javier Sánchez-Hernández[2], Michael E. McClain[3], Rashid Tamatamah[4], Shelard Chilemeji Mukama[5], Francisco Martínez-Capel[1].

[1]Institut d'Investigació per a la Gestió Integrada de Zones Costaneres (IGIC), Universitat Politècnica de València, Paranimf 1, 46730 Grau de Gandia (València), País Valencià, Spain.

[2] Departamento de Zooloxía, Xenética e Antropoloxía Física, Universidade de Santiago de Compostela, Campus Vida s/n, 15782 Santiago de Compostela, Spain.

[3]UNESCO-IHE Institute of Water Education 2611 DA, Delft, The Netherlands.

[4]Department of Aquatic Sciences and Fisheries University of Dar es Salaam P.O. Box 35064, Dar es Salaam, Tanzania.

[5]Department of Environmental Studies, Faculty of Sciences, Technology and Environmental Studies P. O. Box 23409, Dar es Salaam, Tanzania.

Abstract

The influences of habitat structure and hydraulics on tropical macroinvertebrate communities were investigated in two foothill rivers of the Udzungwa Mountains (United Republic of Tanzania) to assist future Environmental Flow Assessments (EFAs). Macroinvertebrate samples, hydraulic variables and habitat structure were collected at the microhabitat scale ($n = 90$). Macroinvertebrate communities were first delineated (i.e. clustered) through Poisson and negative binomial mixture models for count data in a semi-supervised mode by taking into account the sampled river. Then, genetically optimised Multi-Layer Perceptrons (MLPs) were used to identify the relationship of the most relevant variables with the delineated communities. Between the three delineated communities exclusively one community was shared between both rivers. The first and third communities presented similar values of richness (i.e. number of families) and diversity but the first was characterised by high abundance and was dominated by Baetidae (43.2%) while Hydropsychidae (36.3%) dominated the third community. The second community was dominated by Baetidae (33.4%), but it involved low abundance, richness and diversity samples and encompassed the microhabitats where no-macroinvertebrates were found. The performance of the MLP acknowledged the quality of the delineation and it indicated that the first community shows a clear affinity for microhabitats with aquatic vegetation and woody debris and the third for unshaded, fast flowing and shallow microhabitats on intermediate-sized substrate. Conversely, the second community occurred in deep and shaded microhabitats with low flow velocity and coarse substrate. These results should enhance the implementation of ongoing and future EFA studies.

Keywords

Africa; Artificial neural network; Community ecology; Count data; Environmental flow assessment; Semi-supervised clustering

1 Introduction

The recognition of deleterious human activities on freshwater ecosystems is well recognised (Zalewski, 2008). For instance, the construction of infrastructure to guarantee water supply for humans has led to anthropogenic effects through flow alteration and regulation (Kundzewicz, 2007). These negative impacts spreading rapidly in developing tropical and sub-tropical countries, where the urgent need to use water for economic development overrides the implementation of initiatives promoting environmental protection (Msuya and Lalika, 2017). Environmental protection can be accomplished through specific actions on living organisms and habitat conservation. Concerning riverine habitats, the core importance of habitat structure and hydraulics are well recognised (Clifford et al., 2006 and references therein), and hydrology has been considered as a key variable affecting the dynamics and distribution patterns of freshwater species populations (see e.g. Schiemer, 2016). In this context, Environmental Flow Assessment (EFA) has emerged as a fundamental tool to determine the quantities, quality, and patterns of water flows (i.e. environmental flows or e-flows) to balance the protection of the natural environment with out-of-stream uses (McClain et al., 2013). Between the different approaches to EFA, the scientific community currently advocates holistic approaches, which consider the different components (e.g. riparian vegetation, macroinvertebrate communities and fish assemblages) and processes (e.g. matter fluxes) of riverine and riparian ecosystems and account also for human needs.

Among these components, benthic macroinvertebrates are considered as one of the most relevant taxa to assess the ecological integrity of aquatic ecosystems (e.g. Park et al., 2003). Macroinvertebrates are ubiquitous, largely dependent on the aquatic environment and are especially sensitive to flow and stream temperature changes (White et al., 2017 and references therein). Therefore, understanding how communities can change with respect to environmental variables (i.e. flow and eco-hydraulic relationships) is a fundamental basis for ecosystem management and EFA (Belmar et al., 2013). In this regard, clustering techniques can be useful to delineate communities to serve as targets to develop the necessary eco-hydraulic relationships (Adriaenssens et al., 2007). In accordance, these relationships have been typically addressed following two-step approaches: first communities are delineated (i.e. clustered) and then, relationships are inferred (Park et al., 2003). Unfortunately, the former task is not easy because over-dispersion and nonlinear-complex interactions occur in datasets consisting of many species and sampling areas (Adriaenssens et al., 2007; Park et al., 2003).

The aforementioned interactions and nonlinearity triggered the popularity of several sophisticated statistical and machine learning approaches. For instance, a common technique employed to delineate macroinvertebrate communities is Self-Organizing Maps (SOMs) (Kohonen, 1982), which is a kind of artificial neural network (Adriaenssens et al., 2007; Park et al., 2003; Song et al., 2006). However, SOMs and many other technics require data standardisation – because they are sensitive to data over-dispersion (e.g. Song et al., 2006; Adriaenssens et al., 2007) – which may ultimately determine the taxa included within each delineated community (Thorne et al., 1999). In this regard, novel clustering approaches particularly designed to handle count data and over-dispersion, such as Poisson or negative binomial mixture models (Si et al., 2014), should be particularly well suited to delineate macroinvertebrates communities.

Despite the aforementioned advances in the analysis of macroinvertebrate communities, studies in tropical rivers, especially on African streams and rivers, have followed more traditional approaches, such as non-metric multi-dimensional scaling (e.g. Baker et al., 2016; Dallas, 2004; Niba and Mafereka, 2015) or several variants related to correspondence and redundancy analysis (e.g. Kasangaki et al., 2006; Chakona et al., 2009). Additionally, the majority of these studies characterising several macroinvertebrate-environment relationships have mainly focused on water quality (e.g. Chakona et al., 2009; Shimba and Jonah, 2016) and land use changes (i.e. natural-forested *vs.* altered-agricultural) (e.g. Kasangaki et al., 2008; Chakona et al., 2009), whereas hydrologic and hydraulic variables have been used less often and exclusively in combination with other environmental predictors (e.g. Kasangaki et al., 2006; Watson and Dallas, 2013). Small-scale differences in hydraulic conditions characterised by water velocity, depth and substrate roughness are useful to predict the spatial distribution of macroinvertebrate assemblages (Brooks et al., 2005). In accordance, eco-hydraulic relationships based on macroinvertebrate communities collected at small spatial scales can be fundamental for EFA (Song et al., 2006). Regrettably, the majority of studies that differentiated spatial scales have focused on comparing reach-scale and basin-scale features (e.g. Minaya et al., 2013). Thus, specific studies focuses on these small spatial scales have not been addressed in most territories, although some have incidentally found relevant differences at sub-reach-scales (Mathooko, 2001; Niba and Mafereka, 2015) highlighting the importance of the patch scale to detect macroinvertebrate variation (Boyero and Bosch, 2004). That said, we still lack a comprehensive understanding of methods to study EFAs and animal communities at small (i.e. microhabitat) scales.

In order to improve our knowledge and provide guidelines for adequate EFAs, this study investigated the role of habitat structure and hydraulics, at the microhabitat scale, on tropical macroinvertebrate communities in two tributaries of the Kilombero River located in the foothills of the Udzungwa Mountains (United Republic of Tanzania). To achieve this aim, (i) the communities were delineated (i.e. clustered) by means of Poisson and negative binomial mixture models in a semi-supervised mode by taking into account the sampled river and (ii) the most relevant variables, and the relationship of these variables with the delineated communities, were sought with genetically optimised artificial neural networks. Finally, the community preferences and the implications for EFA were discussed for application in further studies.

2 Materials and Methods

2.1 Study area

The Kilombero River Basin is characterised by a sub-humid tropical climate with relative humidity ranging from 70 to 80% with an annual rainfall of about 1200 to 1400 mm and two rainy seasons: a long rainy season in March to May and a shorter one around October to December (Mombo et al., 2011). Temperatures normally vary from 20 to 30 °C (Mombo et al., 2011). Human-related activities such as overgrazing by livestock, agriculture and human settlement are threatening the Kilombero basin (Elisa et al., 2010). The data were collected to evaluate lower flows (i.e. after water abstraction). In accordance, the survey was undertaken during one week in the end of January 2015 (i.e. short dry season preceding the long rainy season). During that and the preceding weeks no higher flows occurred.

The sampled rivers were the Udagaji and Mgugwe, which are two small unregulated rivers that flow southwards from the Udzungwa Mountains National Park (Fig. 1). The Udagaji catchment is densely forested whereas the Mgugwe catchment is covered by forest and shrubs in similar proportions. Although the Udagaji River has been identified as possible water source for a large irrigation scheme in the Kilombero Valley (see O'Keeffe et al., 2017), the basin area of the Mgugwe River is larger (213 *vs.* 25 km^2). In accordance, the mean annual flow of the Mgugwe River corresponds to 2.83 m^3/s (1957-1991) whereas that of the Udagaji River corresponds to 0.81 m^3/s (1957-1991). The maximum and minimum elevation of both sampled rivers did not differ significantly (300/325 and 1637/1802 m a.s.l., respectively) but the mean slope of the Udagaji River is more pronounced (20.2º *vs.* 16.3º in Mgugwe River), causing a flashier flow regime.

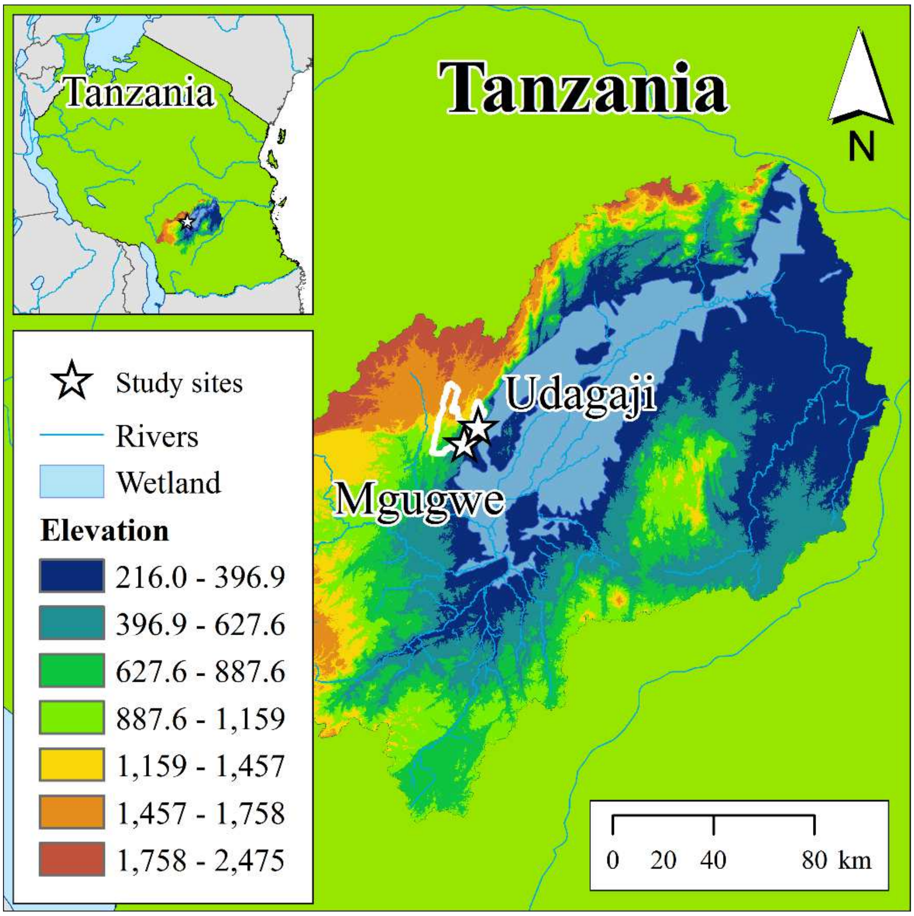

Fig. 1. Location of the Udagaji and Mgugwe rivers and the Kilombero River Basin within the United Republic of Tanzania.

2.2 Data collection

Macroinvertebrate samples were collected at the microhabitat scale – a subset of a mesohabitat (e.g. pool or riffle) defining the homogeneous spatial attributes (e.g. depth, mean column velocity, cover type, and substrate) of physical locations occupied or used by a life stage of a target species or community sometime during its life cycle (*sensu* Bovee et al., 1998). Using the kicking method with a Wildco 500-µm kick net (Yulee, FL, USA), the surveyors quietly moved zigzagging from downstream to upstream sampling systematically the different microhabitats from shore to shore; the distance between microhabitats ranged between 10-15 m. In accordance with the developing plans, the microhabitat preference models were originally intended to evaluate different management scenarios for the Udagaji River. Therefore, the total number of microhabitat replicates sampled (*n* = 90) in the Udagaji River outnumbered those in the Mgugwe River ($n_{Udagaji}$ = 69 and n_{Mgugwe} = 21). In each replicate, three sub-replicates were sampled kicking the substrate for periods of 60 seconds for each replicate (Madikizela and Dye, 2003). After collection, samples were preserved using 70% ethanol and, later in the laboratory, benthic invertebrates were sorted and identified to the family level. No macroinvertebrates were found in 20 microhabitat replicates (13 in the Udagaji River and 7 in the Mgugwe River).

The macroinvertebrate community of each microhabitat replicate (thereafter 'microhabitat') was characterised based on abundance, richness and diversity. Macroinvertebrate abundance was calculated as the total number of individuals per microhabitat (i.e. summing the number of individuals collected in the three replicates). In addition, rarefaction was used to estimate sample richness (i.e. number of families present per microhabitat) and the Shannon–Weiner and Simpson diversity indices, which were calculated using *R* (R Core Team, 2017) package *iNEXT* (Hsieh et al., 2016). These parameters were used to characterise the delineated (i.e. clustered) communities.

Concomitantly to the macroinvertebrate sampling, three hydraulic variables (depth, mean flow velocity and substrate composition) and four factors characterising the structure of the microhabitat (i.e. presence and abundance of reeds, aquatic vegetation, log jams and small woody debris and shade) were measured and scored at three points where each replicate was collected. Later, these values were averaged to define the environmental conditions of each microhabitat.

Depth (m) was measured with a wading rod (to the nearest cm) and the mean flow velocity of the water column – hereafter velocity (m/s) – was measured with a propeller current meter (OTT®) at 40% of the measured depth. The percentage of each substrate class was visually estimated around the sampling point following a simplification of the American Geophysical Union size scale, namely silt (Ø ≤ 62 µm), sand (62 µm > Ø≤ 2 mm), fine gravel (2 > Ø ≤ 8 mm), gravel (8 > Ø ≤ 64 mm), cobbles (64 > Ø ≤ 256 mm), boulders (Ø > 256 mm) and bedrock (Muñoz-Mas et al., 2012). Later, these percentages were aggregated into a single value through the dimensionless substrate index (Mouton et al., 2011). This index is calculated by summing the weighted percentages of each substrate class as follows: $substrate\ index = 0.03 \times Sand\,\% + 0.04 \times Fine\ Gravel\,\% + 0.05 \times Gravel\,\% + 0.06 \times Cobble\,\% + 0.07 \times Boulder\,\% + 0.08 \times Bedrock\,\%$. Finally, the four factors characterising the structure of the microhabitat were scored as absent, scarce, normal or abundant (i.e. from 0 to 3) (Muñoz-Mas et al., 2016b). The microhabitats sampled in the Mgugwe River were deeper and coarser (Fig. 2). In addition, aquatic vegetation was only present in the Mgugwe River.

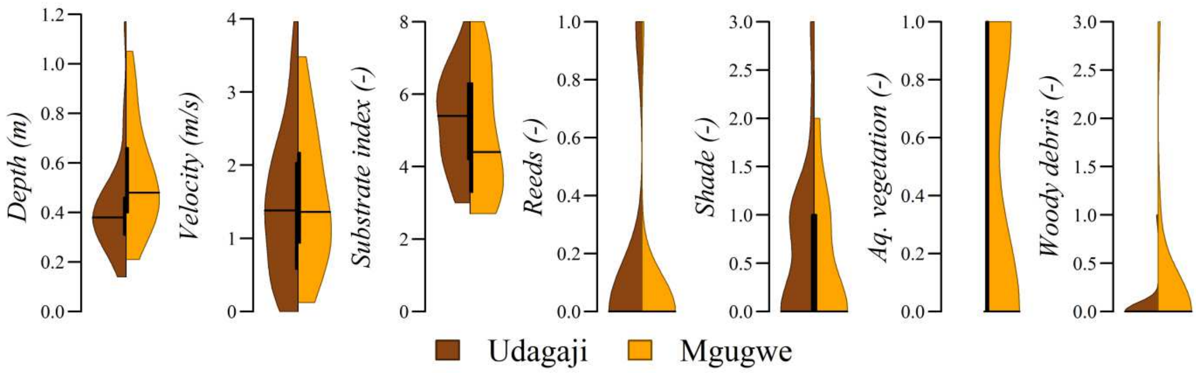

Fig. 2. Violin plots summarising the microhabitat data collected in the Udagaji and Mgugwe rivers (Kilombero River Basin – United Republic of Tanzania). Substrate index, reeds, shade, aquatic vegetation and woody debris are dimensionless.

The force-directed graph (Fruchterman and Reingold, 1991) based on the correlation obtained with the *R* package *polycor* (Fox, 2010), which is specially designed to handle continuous and categorical data, indicated that the hydraulic variables (i.e. depth, velocity and substrate index) were significantly related (Fig. 3). Velocity was positively correlated with substrate, which was negatively correlated with depth. The factors characterising the structure of the microhabitats were not related and neither were with the hydraulic variables, although aquatic vegetation was slightly and positively correlated to velocity.

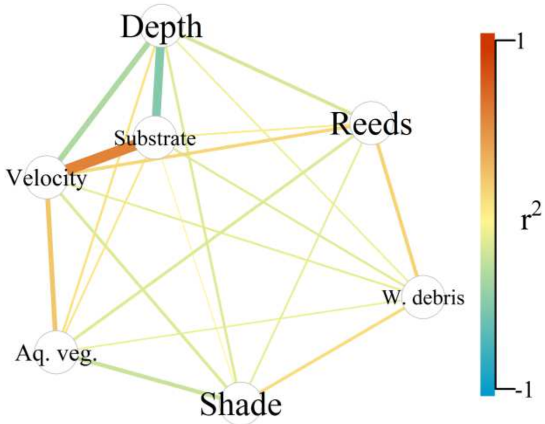

Fig. 3. Force-directed graph based on the correlation (Pearson r^2) between the hydraulic variables and factors collected at each microhabitat obtained with the *R* package *qgraph* (Epskamp et al., 2012).

2.3 Macroinvertebrate community delineation - data clustering

The macroinvertebrate communities present in the foothill rivers of the Udzungwa Mountains were delineated based on the abundance of each family (i.e. number of individuals per family) following the process described in the *R* package *optCluster* (Sekula et al., 2017). This package allows finding the optimal clustering algorithm along with the optimal number of clusters (i.e. communities). In accordance, a number of different approaches with the potential number of communities (i.e. number of clusters) are tested and, for each combination, up to nine validity indices are calculated. There is not a single validity index that outperforms in every situation (Arbelaitz et al., 2013). Therefore, the different combinations are subsequently ranked on the basis of the selected validity indices to obtain the optimal clustering approach and number of clusters (Sekula et al., 2017).

The model-based family of algorithms designed to count data and over-dispersion (i.e. Poisson and negative binomial mixture models) were tested to delineate between 2 and 9 macroinvertebrate communities. Standard model-based clustering algorithms assume that data are generated by a mixture of normal (i.e. Gaussian) probability distributions where each component corresponds to one cluster (Si et al., 2014). However, the macroinvertebrate counts typically involve large numerical differences (i.e. over-dispersion), which compelled scientists to recommend data transformation before clustering (e.g. Adriaenssens et al. 2007). To avoid this step, the tested clustering algorithms – originally included within the *R* package *MBCluster.Seq* (Si, 2012) – employ mixtures of Poisson or negative binomial distributions (Si et al., 2014).

The package *MBCluster.Seq* includes six different variants (three Poisson and three negative binomial alternatives) differing exclusively in the training algorithm used to determine the internal parameters. The first pair is trained with the Expectation Maximization (EM) algorithm (Dempster et al., 1977), which is the most popular method for approximating maximum likelihood estimate (Si, 2012). However, a well-known problem associated with EM is that it can be trapped at local maxima and consequently fails to reach global maxima (Si, 2012). To overcome this limitation, the package *MBCluster.Seq* includes two alternative algorithms, the Simulated Annealing (SA) (Celeux and Govaert, 1992) and Deterministic Annealing (DA) (Rose, 1998).

Although previous studies indicated that differences among environmental conditions (e.g. different depth, substrate composition or water quality) are the real drivers of macroinvertebrate communities (Baker et al., 2016; Costa and Melo, 2008), macroinvertebrate surveys usually collect a limited number of variables, which may limit the predictive capacity of the incomplete variable set. In such a situation, a variable describing the origins of the sample (e.g. sampled river) may be a better predictor because it implicitly encompasses the variables that have not been accounted for, especially when the sampled habitats present evident differences (i.e. depth, substrate and particularly the absence of aquatic vegetation in the Udagaji River). Therefore, although the environmental conditions were not involved in the community delineation, we ranked the different combinations of clustering techniques and number of clusters based on two biological validation indices: the biological homogeneity index (BHI) and the biological stability index (BSI) (Datta and Datta, 2006), which take into account the origins of each sample (i.e. the river where the sample was collected). This semi-supervised approach measures whether, on average, genes (i.e. macroinvertebrate communities sampled in each microhabitat) belonging to the same cluster also belong to the same functional class (i.e. river) (Visconti et al., 2014); but, unlike other semi-supervised methods, it does not enforce or prevent any particular aggregation (Jain, 2010). The BHI evaluates how similar defined clusters are by calculating the average proportion of paired genes (i.e. pair of sampled communities) that are clustered together and have the same functional class (i.e. were collected in the same river). Conversely, the BSI examines the consistency of clustering similar biologically functioning genes together (i.e. belonging to the same river). Observations (i.e. macroinvertebrate families) are removed from the dataset one at a time and the cluster assignments of genes (i.e. sampled communities) with the same functional class (i.e. belonging to the same river) are compared to the cluster assignments based on the full dataset.

The function *repRankAggreg* – originally included within the *R* package *RankAggreg* (Pihur et al., 2009) – was used to infer the optimal clustering algorithm along with the optimal number of clusters. This function performed a weighted rank aggregation of the 6×8 tested combinations following a Monte Carlo cross-entropy approach to render the optimal number of clusters accounting simultaneously and equally for the two validity indices (Pihur et al., 2007).

Finally, the abundance, richness and Shannon–Weiner and Simpson diversity indices of the communities delineated by the optimal clustering approach and number of clusters determined with *repRankAggreg* were compared with the Bayesian test implemented within the *R* package *BEST* (Kruschke, 2013), which provides credible values of the mean, median and standard deviation to infer their differences. The member and counts of each delineated community were inspected and the resulting clusters were used in subsequent analyses.

2.4 Eco-hydraulic relationships inference - neural networks-based classification

The most relevant variables, and the relationship of these variables with the delineated communities (i.e. clusters), were sought with genetically optimised Multi-Layer Perceptrons (MLPs) (McCulloch and Pitts, 1943; Rumelhart et al., 1986). MLPs are a kind of feedforward artificial neural network inspired by the structure of the nervous system with three or more layers of fully-connected neuron-nodes (Olden et al., 2004). Three layered (input-layer, hidden-layer, output-layer) MLPs were developed with the *R* package *nnet* (Venables and Ripley, 2002). The same number of output neurons as the number of delineated communities (i.e. clusters) was used (Walczak and Cerpa, 1999) and the outputs of the linear functions were standardised employing the *softmax* function. This permitted to infer the suitability, between zero and one, of a given microhabitat to each delineated community in a comprehensible manner.

To prevent overfitting, we simultaneously sought the optimal weights for each community, number of neuron nodes and microhabitat variable subset (Goethals et al., 2007). We used a *wrapper* approach involving cross-validation and the Genetic Algorithm (GA) (Holland, 1992) implemented within the *R* package *rgenoud* (Mebane Jr and Sekhon, 2011), which is an approach that proved markedly proficient (see Muñoz-Mas et al., 2016a and therein references) to search them. The optimisation was performed following a repeated k-fold scheme ($10 \times 10_{cross-validation}$), with every fold presenting a similar proportion of samples per community (i.e. samples per cluster) to the original dataset and the performance criterium was the balanced accuracy (i.e. the number of correctly predicted cases weighted by the rarity of the community), which ranges between 0–1 (Muñoz-Mas et al., 2016c). The nine different operators that govern the optimisation performed by the GA (Mebane Jr and Sekhon, 2011) were selected to avoid premature convergence, as previously suggested (Muñoz-Mas et al., 2017). In this study, the population size was set after $N_{population} = 10 \times (N_{clusters} + 1 + N_{predictors})$ and the optimisation halted after a similar number of generations without improvement whereas the maximum number of generations was set to $10 \times N_{population}$.

The variable importance was examined following the Olden approach (Olden et al., 2004), which calculates the importance as the product of the raw input-hidden and hidden-output connection weights between each input and output neuron and sums the product across all hidden neurons (Beck, 2016). The method was implemented using the *R* package *NeuralNetTools* (Beck, 2016) and it was calculated for the 100 MLPs that presented the best generalisation to calculate confidence intervals. Finally, the modelled relationship between the selected variable subset and the probability of presence of each delineated community was graphically characterised with partial dependence plots (Friedman, 2001). Partial dependence plots depict the average of the response variable *vs.* the inspected variable and account for the effects of the remaining variables within the model by averaging their effects. The partial dependence plots were calculated adapting the code appearing in the *R* package *randomForests* (Liaw and Wiener, 2002) and they were likewise calculated for the 100 MLPs that presented the better generalisation to calculate confidence intervals.

3 Results

3.1 Macroinvertebrate communities

A total of 1443 macroinvertebrates were identified. The most abundant order was Ephemeroptera (49.40%), followed by Trichoptera (21.57%) and Lepidoptera (6.39%), whereas the least abundant order was Hemiptera (1.48%). The most abundant families were: Baetidae (28.69%), Hydropsychidae (20.51%) and Leptophlebiidae (14.21%), whereas the least abundant were Tricorythidae (0.07%), Helodidae (0.07%) and Atyidae (0.07%).

Three macroinvertebrate communities were identified (i.e. the optimal number of clusters was three) using the Poisson mixture model trained with DA. Community 1 encompassed 12 samples collected exclusively in the Mgugwe River and the Community 3 included 30 samples collected in the Udagaji River. Community 2 was the only cluster encompassing samples collected in both rivers, although most of them were collected in the Udagaji River (39/9) (Table 1). Community 1 presented higher abundance, although richness and the diversity indices were similar to those of Community 3 (Fig. 4). Conversely, Community 2 presented the lowest values of abundance, richness and the diversity indices.

Table 1. Number of samples per river encompassed within each delineated community.

River/Community	Community 1	Community 2	Community 3
Mgugwe	12	9	0
Udagaji	0	39	30

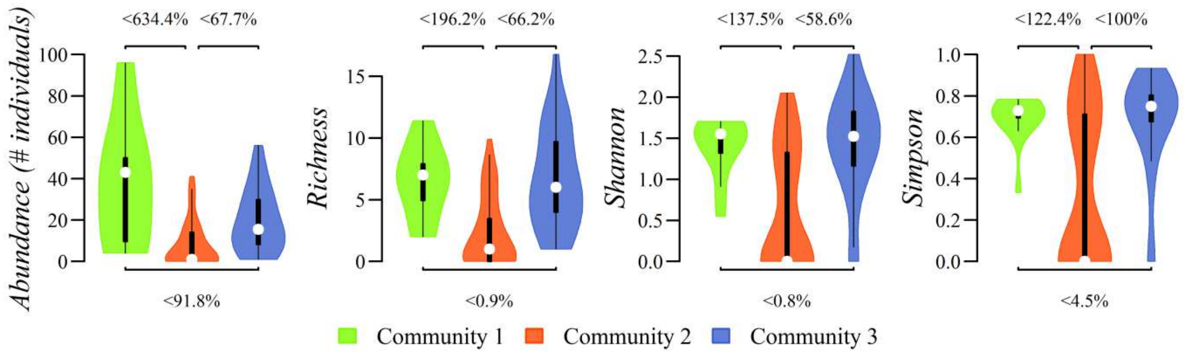

Fig. 4. Violin plots depicting the distribution of the community indices for the three delineated communities; the tagged percentages depict the differences on median values between communities.

The analysis per order and family corroborated the aforementioned general pattern in abundance, although the total number of individuals delineated within Community 3 was higher (Fig. 5). Therefore, the abundance of the samples included within Community 1 (454 ind./12 samples) was higher than in Community 3 (595 ind./30 samples) whereas Community 2 encompassed the least abundant samples (374 ind./48 samples).

Between communities, the most abundant families in Community 1 were Baetidae (43.17%), Pyralidae (20.04%) and Hydropsychidae (10.79%), whereas Hydropsychidae (36.30%), Leptophlebiidae (15.63%), Baetidae (15.63%) and Potamonautidae (12.27%) were the most abundant in Community 3. Conversely, Community 2 was dominated by Baetidae (33.42%), Leptophlebiidae (29.95%) and Perlidae (10.43%); the empty microhabitats (i.e. the 20 microhabitats without macroinvertebrates) were aggregated to Community 2.

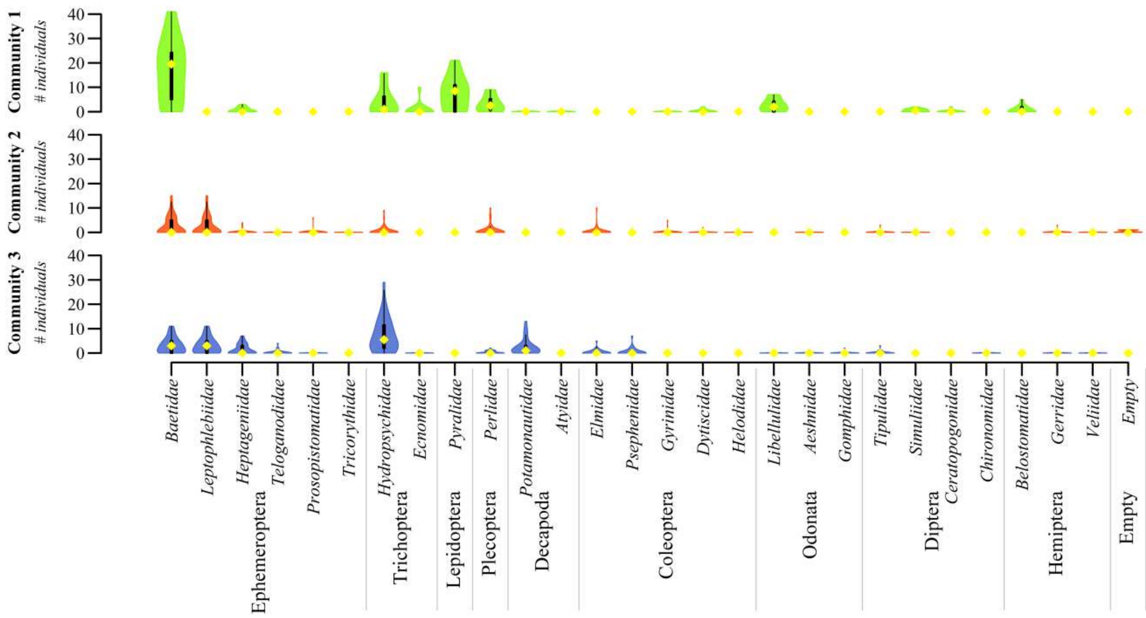

Fig. 5. Violin plots depicting the distribution of the abundance (# of individuals) of each family within the three delineated communities. The families are sorted first by order abundance and then by family abundance.

3.2 Eco-hydraulic relationships

The MLP structure that generalised most over the validation datasets was obtained with three neuron-nodes in the hidden-layer and overweighing Community 2 (57.07%) compared to the other two communities (Community 1 = 21.33% and Community 3 = 21.60%). The better performance was obtained with six variables, namely depth, velocity, substrate index, shade, aquatic vegetation and woody debris and the mean balanced accuracy per community achieved very high values (i.e. Community 1 = 0.84±0.21, Community 2 = 0.77±0.12 and Community 3 = 0.84±0.11).

The partial dependence plots indicated that Community 1 had a clear affinity for microhabitats with aquatic vegetation and woody debris and, to a lesser extent, for finer substrates (i.e. sands) (Fig. 6). Community 2 occurred in deep and shaded microhabitats with low flow velocity and the coarsest substrates (including bedrock). Conversely, Community 3 occurred in unshaded, shallow fast flowing microhabitats with intermediate substrate (i.e. gravel and fine gravel).

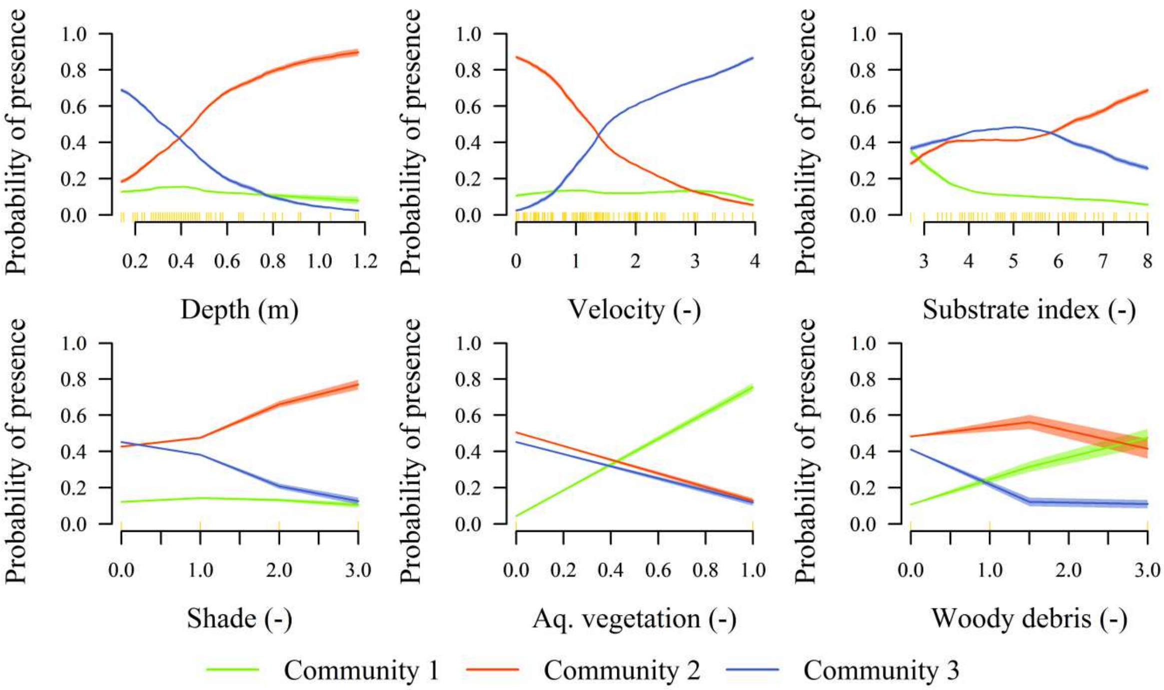

Fig. 6. Mean partial dependence plots, and confidence interval, of the six selected variables. These plots depict the relationship between each variable and the probability of presence of the three delineated communities.

The variable importance analysis corroborated the trends observed in the partial dependence plots, with aquatic vegetation and woody debris, followed by velocity, as the most discriminant variables for Community 1 (Fig. 7). These three variables were likewise the most important for Community 2, although they presented the opposite effect (i.e. sign). Finally, the most important variables for Community 3 were velocity, depth and substrate.

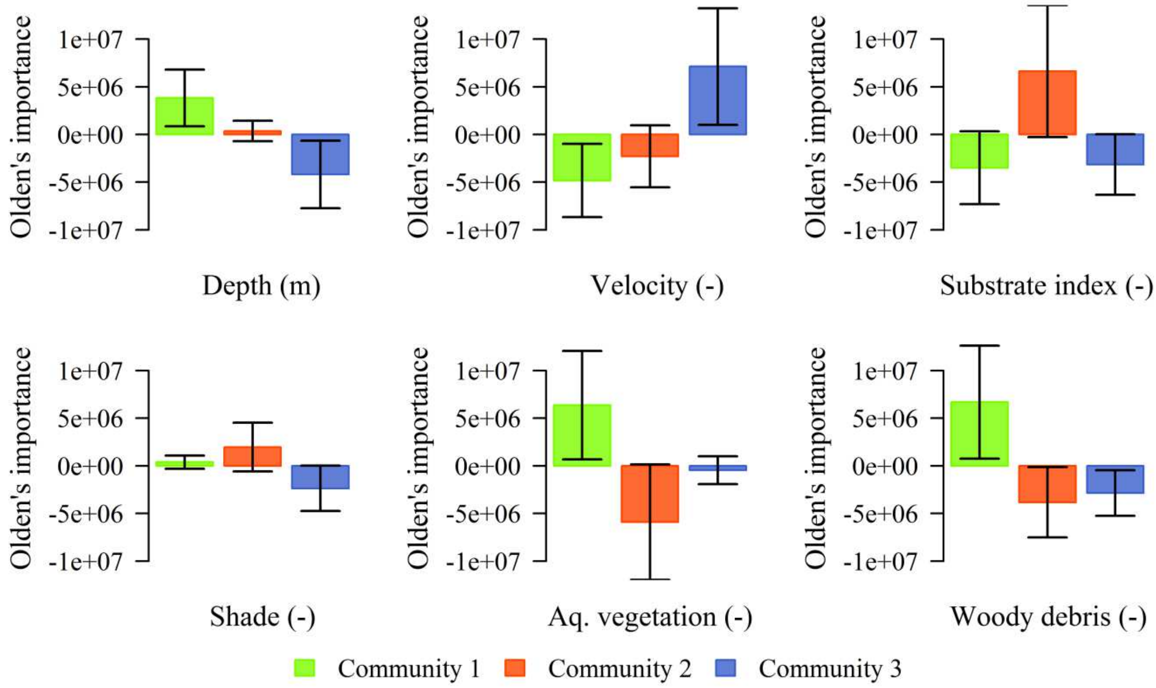

Fig. 7. Variable importance computed with the Olden approach (Olden et al., 2004) for the three delineated communities.

4 Discussion

A central challenge in community ecology is to understand the mechanisms that shape animal assemblages. Our study corroborated that habitat structure and hydraulics also play a fundamental role in shaping the macroinvertebrate communities in the foothill rivers of the Udzungwa Mountains (Baker et al., 2016; Costa and Melo, 2008). We demonstrated that habitat structure and hydraulics are able to properly discriminate the macroinvertebrate communities, which, in turn, underlines their importance as drivers of community composition and abundance. Aquatic vegetation, woody debris, velocity and substrate index, followed by depth and shade, emerged as the most discriminant variables to understand macroinvertebrate communities in these tropical running waters.

4.1 Macroinvertebrate communities

We demonstrated that the optimal number of communities and clustering algorithm can be found with the functionalities implemented within the *optCluster* (Sekula et al., 2017), which allowed us to determine three types of macroinvertebrate communities in a semi-supervised mode by taking into account the sampled river. We indicated that exclusively one community was shared between both rivers. The quality of the aggregation is acknowledged by the results obtained with the MLP, which achieved very high performance (mean balanced accuracy ≈ 0.80). Compared to previous studies (e.g. Park et al., 2003; Edia et al., 2010), the MLP presented in this study performed well with three neuron-nodes and six variables, although former studies did not apply exactly the same approach followed here. Furthermore, the number of delineated communities (i.e. three) was in line with other studies that used SOM in a similar manner (e.g. Park et al., 2003; Edia et al., 2010). In accordance, the use of model-based clustering algorithms assuming that data were generated by a mixture of Poisson or negative binomial probability distributions following semi-supervised mode approaches should be taken into account as a general framework in further studies pooling data from different river segments (Si et al., 2014).

Concerning to the macroinvertebrate composition, the most abundant family was Baetidae, which is globally distributed (Dallas, 2004; Mathooko and Mavuti, 1992), and thus it cannot be considered particularly indicative, although its low abundance has been stated to be indicative of impoverished ecological status (Elias et al., 2014; Shimba and Jonah, 2016; Zhang et al., 2018). Another widely distributed taxa, Diptera, was not abundant compared to the reference sites sampled in other studies focused on African systems (Dallas and Mosepele, 2007; Kasangaki et al., 2006; Mathooko and Mavuti, 1992). Therefore, the largest differences between the macroinvertebrate communities of the Udzungwa Mountains and those sampled in other studies were found for river stretches sampled in the vicinity of large populations; where the water quality led to markedly different communities dominated by individuals of the order Diptera (Elias et al., 2014; Shimba and Jonah, 2016). Although the composition of the macroinvertebrate communities may remain markedly constant (Dallas, 2004; McClain et al., 2014), care must be taken in interpreting these results in terms of abundance because changes in composition may be governed by small and temporary changes (McClain et al., 2014).

4.2 Eco-hydraulic relationships

We identified aquatic vegetation, woody debris, velocity, substrate index and, to a lesser extent, depth and shade as the most discriminant variables to understand macroinvertebrate communities in the studied tropical rivers. In the past, the use of depth and velocity and not the combined effect in the form of shear stress or Froude number has been criticised (Mérigoux et al., 2009). However, the best MLP was obtained employing simultaneously velocity, substrate index and depth and considering fully interacting variables, which has been suggested to increase predictive capacity (Mérigoux et al., 2009). With this variable set, the MLP achieved very high performance and led us to consider the use of these derived variables potentially redundant. Former studies faced difficulties to distinguish macroinvertebrate communities (Adriaenssens et al., 2007) while our results found a clear separation for the three delineated communities according to key environmental variables (here aquatic vegetation and substrate index). Nevertheless, the relative narrow spectrum of sampled conditions may have favoured a better discrimination than other studies that encompassed a larger variability and worked at a lower taxonomic level (i.e. species level) (e.g. Adriaenssens et al., 2007; Mérigoux et al., 2009), especially taking into account that in our case several families appeared spread over different communities.

Interestingly the most relevant variables, and their impact on macroinvertebrate abundance and composition, fit well with *a priori* classifications performed in other studies where the available habitats were classified as stones, vegetation or sand accounting for the type (bedrock rapid *vs.* cobble riffle) and quality (deposition of silt on stones) of the underlying substrate (Dallas, 2007). These differences between vegetated *vs.* non-vegetated and sandy *vs.* coarse substrate have been reported in other African streams, most likely because some of them compared to others are complex habitats that provide (i) refuge from current and fish predation, (ii) food supply for herbivores and detritivores, (iii) attachment for filter-feeding taxa and (iv) exit points for emerging aquatic insects (Chakona et al., 2008). In particular, macrophytes enhance the physical and chemical heterogeneity in aquatic ecosystems (Phiri et al., 2011), and density increases of vegetation have been related with changes in invertebrate body size distribution, with large-bodied individuals and taxa generally being more abundant in dense vegetation owing to the reduction in predation efficiency and foraging success of fish (Phiri et al., 2011). Thus, our outcomes are in agreement with these considerations highlighting the key importance of aquatic vegetation in the structure of macroinvertebrate communities.

Similar reasoning can be applied to woody debris because Ephemeroptera and Trichoptera often feed on leaf litter and/or hide in woody debris (Cummins and Klug, 1979). Usually, the presence of woody debris is particularly relevant at least for some Trichoptera because it provides the necessary material to build their characteristic cases (de Moor and Ivanov, 2008). However, this might not be the case in this study as the identified Trichoptera (Hydropsychidae and Ecnomidae) are caseless (de Moor, 2005). Still, small woody debris can be of importance to aquatic invertebrates as, for instance, a food source for many species (e.g. Cummins and Klug, 1979). Although it may be not exempt from controversy (Aguiar et al., 2017; Lau et al., 2008), it has been stated that in African rivers deforestation and bank-cultivation, and the consequent reduction in the income material, are a main cause of their absence (Chakona et al., 2009).

The importance of velocity, substrate and depth, which presented the most significant correlations (Fig. 3), has been highlighted in a number of studies performed in tropical rivers either on the African continent (Chakona et al., 2009; Dallas, 2007) or in other tropical regions (Baker et al., 2016; Boyero and Bosch, 2004). Nonetheless, habitats with the same substrate composition but different flow velocity or depth often harbour different macroinvertebrate communities (Bauernfeind and Moog, 2000). Setting aside the results obtained for microhabitats with aquatic vegetation, which may mask the effect of the hydraulic variables, the correlation between velocity and substrate observed in this study support the view of former studies suggesting that Ephemeroptera and Trichoptera prefer to inhabit riffle type habitats with coarse substrate (Bauernfeind and Moog, 2000; Chakona et al., 2009; Mathooko, 2001) because these two orders were abundant in Community 3. However, they were also significantly abundant – especially Baetidae (Ephemeroptera) – in Community 1, which was related to sandy substrate. Sandy substrates are usually unstable and disfavour macroinvertebrate settlement (Duan et al., 2009). Therefore, we hypothesise that microhabitats dominated by sandy substrate, which presented communities that usually occur in riffles (Duan et al., 2009), were in general near the banks and subject to lower stresses. Therefore, this spatial distribution may have favoured the establishment of aquatic vegetation where they feed and find protection from predators, which permits their proliferation (Masese et al., 2014) and thus, substrate was in this case of minor relevance. In contrast, the result obtained for the coarsest substrate (i.e. bedrock) does not pose any doubt because this substrate usually renders little space for the macroinvertebrate refuge (e.g. holes or crevices), which justifies the impoverished communities found over there (Baker et al., 2016).

Perhaps the most contradicting pattern was that related to water depth because previous studies performed in other African streams found a positive effect on macroinvertebrate abundance, particularly on the Ephemeroptera and Trichoptera orders (e.g. Chakona et al., 2009; Masese et al., 2014). Nevertheless, our results accept the view that pools host impoverished macroinvertebrate communities compared to shallower mesohabitats (e.g. riffles) as observed in other tropical streams (Baker et al., 2016). We posit that this discrepancy may be caused by the different scales employed in these studies compared to our study, which was performed at the microhabitat scale and encompassed relatively short river segments, whereas the discrepant studies were performed at the mesohabitat scale encompassing long river segments that lead to a gradient of depth of different nature.

Unlike temperate rivers, in tropical rivers there is certain controversy about the origin of the primary resources with several authors claiming autochthonous (e.g. periphytic algae and/or cyanobacteria) prevailing over allochthonous origins (e.g. leaf litter) (e.g. Lau et al., 2008) and others claiming the opposite (e.g. Aguiar et al., 2017). The results obtained for shade may indicate that the Udagaji and Mgugwe rivers rely on autochthonous production, although this cannot be considered a general pattern unequivocally transferable to other African rivers (see e.g. Masese et al., 2014). Nonetheless, in other tropical streams density and richness were higher when canopy cover was more variable (Boyero and Bosch, 2004). In accordance, specific research should be performed to elucidate the real causes of such macroinvertebrate distribution patterns in relation to shade.

4.3 Potential implications of altered hydraulics and flow regimes

A common practise worldwide is the construction of infrastructure to guarantee irrigation schemes and water supply for humans with concomitant significant reductions and alterations in river flows. The studied rivers represent systems with natural flow conditions in which no regulatory facilities are planned, but the alteration of hydraulics through irrigation schemes would drive deleterious changes in macroinvertebrate communities and linked components of river food webs. Invertebrate abundance may vary in response to decreased flow, whereas invertebrate richness commonly decreases along with habitat diversity (Boyero and Bosch, 2004; Masese et al., 2013). In this regard, and based exclusively in our results, reductions in river flows and depth that favour the proliferation of macrophytes (Schoelynck et al., 2018) are likely to increase the areas suitable for the community delineated in Community 1, although it may not occur in the Udagaji River. However, the consequent reduction in flow velocity in the downstream reach may negatively impact Community 3, which also presented high richness and diversity. Consequently, although the ultimate impact of water abstraction is rather uncertain, we consider that reductions of river flows caused by water diversion are likely to reduce the overall abundance of macroinvertebrates as has been demonstrated in other streams of south-eastern Africa that suffered significant reductions in flows (Chakona et al., 2008; Mathooko and Mavuti, 1992). That said, large irrigation schemes would modify the geomorphology of the streams and the input of woody material into the river system, which is likely to impact directly shredder species and indirectly other macroinvertebrates or trophic levels through cascading effects (Chakona et al., 2009; Kasangaki et al., 2006). However, the mechanism triggering cascading effects might change among rivers as our results also indicated that shade may be linked to autochthonous primary production through grazing (i.e. scrapers). Small impoundments can withhold sediments, organic debris, and nutrients (Mbaka and Wanjiru Mwaniki, 2015), which will expose downstream river segments to a sediment deficit – fine sediment is likely to flow preferentially trough the irrigation canal with coarser sediment trapped at the point of water diversion (Taniwaki et al., 2017). The upstream river segments will be, on the contrary, negatively impacted by the increased depth caused by the impoundment, which is likely to lead to the impoverished macroinvertebrate communities delineated in Community 2. Although, it is difficult to predict how most species will respond to new environmental conditions, we conclude that water abstraction is unlikely to have neutral effect over the macroinvertebrate communities of the Udagaji and Mgugwe rivers and therefore these practices are not recommended from an ecological conservation perspective.

This study has not been exhaustive and has neglected some physical and chemical variables. In accordance, the ultimate type and magnitude of impacts corresponds to complex interactions that would be observed in the long term (Mbaka and Wanjiru Mwaniki, 2015). Despite increasing concern about how climate and land-use change and river regulation will affect freshwater ecosystems, comparatively a few studies have focused on small tropical streams (Taniwaki et al., 2017). Therefore, the herein presented results provide valuable information on macroinvertebrate communities and eco-hydrological relationships in tropical streams of East Africa, which should adequately guide further ecological studies and assist EFAs.

5 References

Adriaenssens, V., Verdonschot, P.F.M., Goethals, P.L.M., De Pauw, N., 2007. Application of clustering techniques for the characterization of macroinvertebrate communities to support river restoration management. Aquat. Ecol. 41, 387–398. doi:10.1007/s10452-005-2836-0

Aguiar, A.C.F., Neres-Lima, V., Moulton, T.P., 2017. Relationships of shredders, leaf processing and organic matter along a canopy cover gradient in tropical streams. J. Limnol. Vol 77, No 1. doi:10.4081/jlimnol.2017.1684

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J.M., Perona, I., 2013. An extensive comparative study of cluster validity indices. Pattern Recognit. 46, 243–256. doi:10.1016/j.patcog.2012.07.021

Baker, K., Chadwick, M.A., Kahar, R., Sulaiman, Z.H., Wahab, R.A., 2016. Fluvial biotopes influence macroinvertebrate biodiversity in South□East Asian tropical streams. Ecosphere 7, e01479. doi:10.1002/ecs2.1479

Bauernfeind, E., Moog, O., 2000. Mayflies (Insecta: Ephemeroptera) and the assessment of ecological integrity : a methodological approach. Hydrobiologia 422, 71–83. doi:10.1023/A:1017090504518

Beck, M., 2016. NeuralNetTools: Visualization and Analysis Tools for Neural Networks.

Belmar, O., Velasco, J., Gutiérrez-Cánovas, C., Mellado-Díaz, A., Millán, A., Wood, P.J., 2013. The influence of natural flow regimes on macroinvertebrate assemblages in a semiarid Mediterranean basin. Ecohydrology 6, 363–379. doi:10.1002/eco.1274

Bovee, K.D., Lamb, B.L., Bartholow, J.M., Stalnaker, C.B., Taylor, J., Henriksen, J., 1998. Stream habitat analysis using the instream flow incremental methodology. U.S. Geological Survey Information and Technology, Fort Collins, CO (USA).

Boyero, L., Bosch, J., 2004. The Effect of Riffle-Scale Environmental Variability on Macroinvertebrate Assemblages in a Tropical Stream. Hydrobiologia 524, 125–132. doi:10.1023/B:HYDR.0000036127.94781.3c

Brooks, A.J.J., Haeusler, T., Reinfelds, I., Williams, S., 2005. Hydraulic microhabitats and the distribution of macroinvertebrate assemblages in riffles. Freshw. Biol. 50, 331–344. doi:10.1111/j.1365-2427.2004.01322.x

Celeux, G., Govaert, G., 1992. A classification EM algorithm for clustering and two stochastic versions. Comput. Stat. Data Anal. 14, 315–332. doi:10.1016/0167-9473(92)90042-E

Chakona, A., Phiri, C., Day, J.A., 2009. Potential for Trichoptera communities as biological indicators of morphological degradation in riverine systems. Hydrobiologia 621, 155–167. doi:10.1007/s10750-008-9638-z

Chakona, A., Phiri, C., Magadza, C.H.D., Brendonck, L., 2008. The influence of habitat structure and flow permanence on macroinvertebrate assemblages in temporary rivers in northwestern Zimbabwe. Hydrobiologia 607, 199–209. doi:10.1007/s10750-008-9391-3

Clifford, N.J., Harmar, O.P., Harvey, G., Petts, G.E., 2006. Physical habitat, eco-hydraulics and river design: a review and re-evaluation of some popular concepts and methods. Aquat. Conserv. Mar. Freshw. Ecosyst. 16, 389–408. doi:10.1002/aqc.736

Costa, S.S., Melo, A.S., 2008. Beta diversity in stream macroinvertebrate assemblages: among-site and among-microhabitat components. Hydrobiologia 598, 131–138. doi:10.1007/s10750-007-9145-7

Cummins, K.W., Klug, M.J., 1979. Feeding Ecology of Stream Invertebrates. Annu. Rev. Ecol. Syst. 10, 147–172. doi:10.1146/annurev.es.10.110179.001051

Dallas, H.F., 2007. The effect of biotope-specific sampling for aquatic macroinvertebrates on reference site classification and the identification of environmental predictors in Mpumalanga, South Africa. African J. Aquat. Sci. 32, 165–173. doi:10.2989/AJAS.2007.32.2.8.205

Dallas, H.F., 2004. Seasonal variability of macroinvertebrate assemblages in two regions of South Africa: implications for aquatic bioassessment. African J. Aquat. Sci. 29, 173–184. doi:10.2989/16085910409503808

Dallas, H.F., Mosepele, B., 2007. A preliminary survey and analysis of the spatial distribution of aquatic invertebrates in the Okavango Delta, Botswana. African J. Aquat. Sci. 32, 1–11. doi:10.2989/AJAS.2007.32.1.1.138

Datta, S., Datta, S., 2006. Methods for evaluating clustering algorithms for gene expression data using a reference set of functional classes. BMC Bioinformatics 7, 397. doi:10.1186/1471-2105-7-397

de Moor, F.C., 2005. Variation in case construction of Trichoptera larvae in southern Africa, in: Tanida, K., Rossiter, A. (Eds.), Proceedings of the 11th International Symposium on Trichoptera, Osaka. Tokai University Press, Kanagawa (Japan), Osaka (Japan), pp. 107–114.

de Moor, F.C., Ivanov, V.D., 2008. Global diversity of caddisflies (Trichoptera: Insecta) in freshwater. Hydrobiologia 595, 393–407. doi:10.1007/s10750-007-9113-2

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum Likelihood from Incomplete Data via the

EM Algorithm. J. R. Stat. Soc. Ser. B 39, 1–38.

Duan, X., Wang, Z., Xu, M., Zhang, K., 2009. Effect of streambed sediment on benthic ecology. Int. J. Sediment Res. 24, 325–338. doi:10.1016/S1001-6279(10)60007-8

Edia, E.O., Gevrey, M., Ouattara, A., Brosse, S., Gourène, G., Lek, S., 2010. Patterning and predicting aquatic insect richness in four West-African coastal rivers using artificial neural networks. Knowl. Manag. Aquat. Ecosyst. 06p1-06p15. doi:10.1051/kmae/2010029

Elias, J.D., Ijumba, J.N., Mgaya, Y.D., Mamboya, F.A., 2014. Study on Freshwater Macroinvertebrates of Some Tanzanian Rivers as a Basis for Developing Biomonitoring Index for Assessing Pollution in Tropical African Regions. J. Ecosyst. 2014, 1–8. doi:10.1155/2014/985389

Elisa, M., Gara, J.I., Wolanski, E., 2010. A review of the water crisis in Tanzania's protected areas, with emphasis on the Katuma River—Lake Rukwa ecosystem. Ecohydrol. Hydrobiol. 10, 153–165. doi:10.2478/v10104-011-0001-z

Epskamp, S., Cramer, A.O.J., Waldorp, L.J., Schmittmann, V.D., Borsboom, D., 2012. qgraph: Network Visualizations of Relationships in Psychometric Data. J. Stat. Softw. 48, 1–18.

Fox, J., 2010. polycor: Polychoric and Polyserial Correlations.

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Ann. Stat. 29, 1189–1232. doi:10.1214/aos/1013203451

Fruchterman, T.M.J., Reingold, E.M., 1991. Graph drawing by force-directed placement. Softw. Pract. Exp. 21, 1129–1164. doi:10.1002/spe.4380211102

Goethals, P.L.M., Dedecker, A.P., Gabriels, W., Lek, S., De Pauw, N., 2007. Applications of artificial neural networks predicting macroinvertebrates in freshwaters. Aquat. Ecol. 41, 491–508.

doi:10.1007/s10452-007-9093-3

Holland, J.H., 1992. Genetic algorithms. Sci. Am. 267, 66–72.

Hsieh, T.C., Ma, K.H., Chao, A., McInerny, G., 2016. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). Methods Ecol. Evol. 7, 1451–1456. doi:10.1111/2041-210X.12613

Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern Recognit. Lett. 31, 651–666. doi:10.1016/j.patrec.2009.09.011

Kasangaki, A., Babaasa, D., Efitre, J., McNeilage, A., Bitariho, R., 2006. Links Between Anthropogenic Perturbations and Benthic Macroinvertebrate Assemblages in Afromontane Forest Streams in Uganda. Hydrobiologia 563, 231–245. doi:10.1007/s10750-005-0009-8

Kasangaki, A., Chapman, L.J., Balirwa, J., 2008. Land use and the ecology of benthic macroinvertebrate assemblages of high-altitude rainforest streams in Uganda. Freshw. Biol. 53, 681–697. doi:10.1111/j.1365-2427.2007.01925.x

Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biol. Cybern. 43, 59–69. doi:10.1007/BF00337288

Kruschke, J.K., 2013. Bayesian estimation supersedes the T test. J. Exp. Psychol. Gen. 142, 573–603. doi:10.1037/a0029177

Kundzewicz, Z.W., 2007. Global freshwater resources for sustainable development. Ecohydrol. Hydrobiol. 7, 125–134. doi:10.1016/S1642-3593(07)70178-7

Lau, D.C.P., Leung, K.M.Y., Dudgeon, D., 2008. What does stable isotope analysis reveal about trophic relationships and the relative importance of allochthonous and autochthonous resources in tropical streams? A synthetic study from Hong Kong. Freshw. Biol. 54, 127–141.

doi:10.1111/j.1365-2427.2008.02099.x

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 3, 18–22.

Madikizela, B.R., Dye, A.H., 2003. Community composition and distribution of macroinvertebrates in the Umzimvubu River, South Africa: a pre-impoundment study. African J. Aquat. Sci. 28, 137–149. doi:10.2989/16085910309503778

Masese, F.O., Kitaka, N., Kipkemboi, J., Gettel, G.M., Irvine, K., McClain, M.E., 2014. Macroinvertebrate functional feeding groups in Kenyan highland streams: evidence for a diverse shredder guild. Freshw. Sci. 33, 435–450. doi:10.1086/675681

Masese, F.O., Omukoto, J.O., Nyakeya, K., 2013. Biomonitoring as a prerequisite for sustainable water resources: a review of current status, opportunities and challenges to scaling up in East Africa. Ecohydrol. Hydrobiol. 13, 173–191. doi:10.1016/j.ecohyd.2013.06.004

Mathooko, J.M., 2001. Temporal and spatial distribution of the baetid Afroptilum sudafricanum in the sediment surface of a tropical stream. Hydrobiologia 443, 1–8. doi:10.1023/A:1017502421985

Mathooko, J.M., Mavuti, K.M., 1992. Composition and seasonality of benthic invertebrates, and drift in the Naro Moru River, Kenya. Hydrobiologia 232, 47–56. doi:10.1007/BF00014611

Mbaka, J.G., Wanjiru Mwaniki, M., 2015. A global review of the downstream effects of small impoundments on stream habitat conditions and macroinvertebrates. Environ. Rev. 23, 257–262. doi:10.1139/er-2014-0080

McClain, M.E., Kashaigili, J.J., Ndomba, P., 2013. Environmental flow assessment as a tool for achieving environmental objectives of African water policy, with examples from East Africa. Int. J. Water Resour. Dev. 29, 650–665. doi:10.1080/07900627.2013.781913

McClain, M.E., Subalusky, A.L., Anderson, E.P., Dessu, S.B., Melesse, A.M., Ndomba, P.M., Mtamba, J.O.D.D., Tamatamah, R.A., Mligo, C., 2014. Comparing flow regime, channel hydraulics, and biological communities to infer flow-ecology relationships in the Mara River of Kenya and Tanzania. Hydrol. Sci. J. 59, 801–819. doi:10.1080/02626667.2013.853121

McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5, 115–133. doi:10.1007/BF02478259

Mebane Jr, W.R., Sekhon, J.S., 2011. Genetic optimization using derivatives: The rgenoud package for R. J. Stat. Softw. 42, 1–26.

Mérigoux, S., Lamouroux, N., Olivier, J.M., Dolédec, S., 2009. Invertebrate hydraulic preferences and predicted impacts of changes in discharge in a large river. Freshw. Biol. 54, 1343–1356. doi:10.1111/j.1365-2427.2008.02160.x

Minaya, V., McClain, M.E., Moog, O., Omengo, F., Singer, G.A., 2013. Scale-dependent effects of rural activities on benthic macroinvertebrates and physico-chemical characteristics in headwater streams of the Mara River, Kenya. Ecol. Indic. 32, 116–122. doi:10.1016/j.ecolind.2013.03.011

Mombo, F.M., Speelman, S., Van Huylenbroeck, G., Hella, J., Pantaleo, M., Moe, S., 2011. Ratification of the Ramsar convention and sustainable wetlands management: situation analysis of the Kilombero Valley wetlands in Tanzania. J. Agric. Ext. Rural Dev. 3, 153–164.

Mouton, A.M., Alcaraz-Hernández, J.D., De Baets, B., Goethals, P.L.M., Martínez-Capel, F., 2011. Data-driven fuzzy habitat suitability models for brown trout in Spanish Mediterranean rivers. Environ. Model. Softw. 26, 615–622. doi:10.1016/j.envsoft.2010.12.001

Msuya, T.S., Lalika, M.C.S., 2017. Linking Ecohydrology and Integrated Water Resources Management: Institutional challenges for water management in the Pangani Basin, Tanzania. Ecohydrol. Hydrobiol. doi:10.1016/j.ecohyd.2017.10.004

Muñoz-Mas, R., Fukuda, S., Vezza, P., Martínez-Capel, F., 2016a. Comparing four methods for decision-tree induction: A case study on the invasive Iberian gudgeon (Gobio lozanoi; Doadrio and Madeira, 2004). Ecol. Inform. 34, 22–34. doi:10.1016/j.ecoinf.2016.04.011

Muñoz-Mas, R., Garófano-Gómez, V., Andrés-Doménech, I., Corenblit, D., Egger, G., Francés, F., Ferreira, M.T., García-Arias, A., Politti, E., Rivaes, R., Rodríguez-González, P.M., Steiger, J., Vallés-Morán, F.J., Martínez-Capel, F., 2017. Exploring the key drivers of riparian woodland successional pathways across three European river reaches. Ecohydrology 10, e1888. doi:10.1002/eco.1888

Muñoz-Mas, R., Martínez-Capel, F., Schneider, M., Mouton, A.M., 2012. Assessment of brown trout habitat suitability in the Jucar River Basin (Spain): Comparison of data-driven approaches with fuzzy-logic models and univariate suitability curves. Sci. Total Environ. 440, 123–131. doi:10.1016/j.scitotenv.2012.07.074

Muñoz-Mas, R., Papadaki, C., Martínez-Capel, F., Zogaris, S., Ntoanidis, L., Dimitriou, E., 2016b. Generalized additive and fuzzy models in environmental flow assessment: A comparison employing the West Balkan trout (Salmo farioides; Karaman, 1938). Ecol. Eng. 91, 365–377. doi:10.1016/j.ecoleng.2016.03.009

Muñoz-Mas, R., Vezza, P., Alcaraz-Hernández, J.D., Martínez-Capel, F., 2016c. Risk of invasion predicted with support vector machines: A case study on northern pike (Esox Lucius, L.) and bleak (Alburnus alburnus, L.). Ecol. Modell. 342, 123–134. doi:10.1016/j.ecolmodel.2016.10.006

Niba, A.S., Mafereka, S.P., 2015. Benthic macroinvertebrate assemblage composition and distribution pattern in the upper Mthatha River, Eastern Cape, South Africa. African J. Aquat. Sci. 40, 133–142. doi:10.2989/16085914.2015.1028323

O'Keeffe, J., Graas, S., Mombo, F., McClain, M., 2017. Stakeholder-enhanced environmental flow assessment: The Rufiji Basin case study in Tanzania. River Res. Appl. 1–9. doi:10.1002/rra.3219

Olden, J.D., Joy, M.K., Death, R.G., 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. Ecol. Modell. 178, 389–397. doi:10.1016/j.ecolmodel.2004.03.013

Park, Y.-S., Céréghino, R., Compin, A., Lek, S., 2003. Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters. Ecol. Modell. 160, 265–280. doi:10.1016/S0304-3800(02)00258-2

Phiri, C., Chakona, A., Day, J.A., 2011. The effect of plant density on epiphytic macroinvertebrates associated with a submerged macrophyte, Lagarosiphon ilicifolius Obermeyer, in Lake Kariba, Zimbabwe. African J. Aquat. Sci. 36, 289–297. doi:10.2989/16085914.2011.636907

Pihur, V., Datta, S., Datta, S., 2009. RankAggreg, an R package for weighted rank aggregation. BMC Bioinformatics 10, 62. doi:10.1186/1471-2105-10-62

Pihur, V., Datta, S., Datta, S., 2007. Weighted rank aggregation of cluster validation measures: a Monte Carlo cross-entropy approach. Bioinformatics 23, 1607–1615.

R Core Team, 2017. R: A language and environment for statistical computing.

Rose, K., 1998. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. Proc. IEEE 86, 2210–2239. doi:10.1109/5.726788

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536. doi:10.1038/323533a0

Schiemer, F., 2016. Building an eco-hydrological framework for the management of large river systems. Ecohydrol. Hydrobiol. 16, 19–25. doi:10.1016/j.ecohyd.2015.07.004

Schoelynck, J., Creëlle, S., Buis, K., De Mulder, T., Emsens, W.-J., Hein, T., Meire, D., Meire, P., Okruszko, T., Preiner, S., Roldan Gonzalez, R., Silinski, A., Temmerman, S., Troch, P., Van Oyen, T., Verschoren, V., Visser, F., Wang, C., Wolters, J.-W., Folkard, A., 2018. What is a macrophyte patch? Patch identification in aquatic ecosystems and guidelines for consistent delineation. Ecohydrol. Hydrobiol. 18, 1–9. doi:10.1016/j.ecohyd.2017.10.005

Sekula, M., Datta, S., and Susmita Datta, 2017. optCluster: Determine Optimal Clustering Algorithm and Number of Clusters.

Shimba, M.J., Jonah, F.E., 2016. Macroinvertebrates as bioindicators of water quality in the Mkondoa River, Tanzania, in an agricultural area. African J. Aquat. Sci. 41, 453–461. doi:10.2989/16085914.2016.1230536

Si, Y., 2012. MBCluster.Seq: Model-Based Clustering for RNA-seq Data.

Si, Y., Liu, P., Li, P., Brutnell, T.P., 2014. Model-based clustering for RNA-seq data. Bioinformatics 30, 197–205. doi:10.1093/bioinformatics/btt632

Song, M.-Y., Park, Y.-S., Kwak, I.-S., Woo, H., Chon, T.-S., 2006. Characterization of benthic macroinvertebrate communities in a restored stream by using self-organizing map. Ecol. Inform. 1, 295–305. doi:10.1016/j.ecoinf.2005.12.001

Taniwaki, R.H., Piggott, J.J., Ferraz, S.F.B., Matthaei, C.D., 2017. Climate change and multiple stressors in small tropical streams. Hydrobiologia 793, 41–53. doi:10.1007/s10750-016-2907-3

Thorne, R.S.J., Williams, W.P., Cao, Y., 1999. The influence of data transformations on biological monitoring studies using macroinvertebrates. Water Res. 33, 343–350. doi:10.1016/S0043-1354(98)00247-4

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, Fourth. ed. Springer-Verlag

New York, New York (USA). doi:10.1007/978-0-387-21706-2

Visconti, A., Cordero, F., Pensa, R.G., 2014. Leveraging additional knowledge to support coherent bicluster discovery in gene expression data. Intell. Data Anal. 18, 837–855.

Walczak, S., Cerpa, N., 1999. Heuristic principles for the design of artificial neural networks. Inf. Softw. Technol. 41, 107–117. doi:10.1016/S0950-5849(98)00116-5

Watson, M., Dallas, H.F., 2013. Bioassessment in ephemeral rivers: constraints and challenges in applying macroinvertebrate sampling protocols. African J. Aquat. Sci. 38, 35–51. doi:10.2989/16085914.2012.742419

White, J.C., Hannah, D.M., House, A., Beatson, S.J. V, Martin, A., Wood, P.J., 2017. Macroinvertebrate responses to flow and stream temperature variability across regulated and non-regulated rivers. Ecohydrology 10, e1773–n/a. doi:10.1002/eco.1773

Zalewski, M., 2008. Rationale for the “Floodplain Declaration” from environmental conservation toward sustainability science. Ecohydrol. Hydrobiol. 8, 107–113. doi:10.2478/v10104-009-0008-x

Zhang, M., Muñoz-Mas, R., Martínez-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. Sci. Total Environ. 634, 749–759. doi:10.1016/j.scitotenv.2018.04.021