Science of the Total Environment 634 (2018) 749–759

Contents lists available at ScienceDirect

Science of the Total Environment

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

Determining the macroinvertebrate community indicators and relevant environmental predictors of the Hun-Tai River Basin (Northeast China): A study based on community patterning

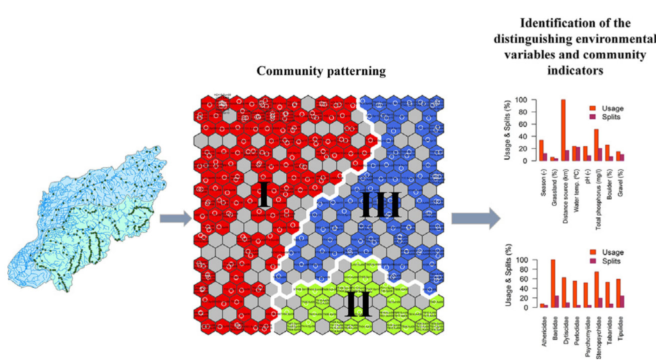
Min Zhang [a,b], Rafael Muñoz-Mas [c], Francisco Martínez-Capel [c], Xiaodong Qu [a,b,*], Haiping Zhang [a,b], Wenqi Peng [a,b], Xiaobo Liu [a,b]

[a] State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100038, China
[b] Department of Water Environment, China Institute of Water Resources and Hydropower Research, Beijing 100038, China
[c] Institut d'Investigació per a la Gestió Integrada de Zones Costaneres (IGIC), Universitat Politècnica de València, C/ Paranimf 1, Grau de Gandia, València 46730, Spain.

HIGHLIGHTS

- SOMs delineated three macroinvertebrate community types in a gradient.
- Local factors affected macroinvertebrate communities under the longitudinal gradient of the geographical features.
- Several indicator families characterized each community type.
- Abundance contributed significantly to distinguish these indicators.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 1 February 2018
Received in revised form 26 March 2018
Accepted 1 April 2018
Available online xxxx

Editor: Henner Hollert

Keywords:
Macroinvertebrate patterning
Indicators
Self-Organizing Map
Decision tree
River zonation

ABSTRACT

It is essential to understand the patterning of biota and environmental influencing factors for proper rehabilitation and management at the river basin scale. The Hun-Tai River Basin was extensively sampled four times for macroinvertebrate community and environmental variables during one year. Self-Organizing Maps (SOMs) were used to reveal the aggregation patterns of the 355 samples. Three community types (*i.e.*, clusters) were found (at the family level) based on the community composition, which showed a clearly gradient by combining them with the representative environmental variables: minimally impacted source area, intermediately anthropogenic impacted sites, and highly anthropogenic impacted downstream area, respectively. This gradient was corroborated by the decreasing trends in density and diversity of macroinvertebrates. Distance from source, total phosphorus and water temperature were identified as the most important variables that distinguished the delineated communities. In addition, the sampling season, substrate type, pH and the percentage of grassland were also identified as relevant variables. These results demonstrated that macroinvertebrates communities are structured in a hierarchical manner where geographic and water quality prevail over temporal (season) and habitat (substrate type) features at the basin scale. In addition, it implied that the local-scale environment variables affected macroinvertebrates under the longitudinal gradient of the geographical and anthropogenic pressure. More than one family was identified as the indicator for each type of community. Abundance contributed significantly for distinguishing the indicators, while Baetidae with higher density indicated minimally and

* Corresponding author at: State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing, 100038, China.
E-mail address: quxiaodong@iwhr.com. (X. Qu).

https://doi.org/10.1016/j.scitotenv.2018.04.021
0048-9697/

intermediately impacted area and lower density indicated highly impacted area. Therefore, we suggested the use of abundance data in community patterning and classification, especially in the identification of the indicator taxa.

1. Introduction

Currently, human activities are greatly influencing the flow rate, water yield, sediment transport and nutrient releases in freshwater ecosystems at scales that far exceed those of natural phenomena (Habersack et al., 2014). Accordingly, water resources are currently over-exploited in many regions, which has resulted in 65% of rivers worldwide being under moderate-to-high threats in terms of human water security and biodiversity loss (Vörösmarty et al., 2010).

Biotic assemblages in freshwater ecosystems integrate these impacts throughout the drainage basins; thus, these assemblages can be considered as indicators of ecosystem health (Habersack et al., 2014). Consequently, the classification and delineation of the ecological statuses of rivers based on the biotic assemblages is an essential prerequisite for river ecosystem assessment, restoration and management (Heino et al., 2002; Marchant et al., 2000; Siddig et al., 2016; Tsai et al., 2017).

Macroinvertebrate assemblages have been widely used as indicators of ecosystem changes because macroinvertebrate communities encompass a diverse group with a wide range of life-history requirements (O'Brien et al., 2016). Macroinvertebrates vary spatially and temporally and integrate ecosystem changes as a result of their suite of feeding strategies and lifestyles and their different sensitivities to changes in physical habitat and water quality (Milošević et al., 2016; Ogbeibu and Oribhabor, 2002). According to a recent review on indicator species over the last 14 years, nearly 50% of the taxa used as indicators were animals, and 70% of these were invertebrates (Siddig et al., 2016). However, data on macroinvertebrate assemblages are highly complex and difficult to analyze because macroinvertebrate assemblages consist of numerous species that respond in complex manners to natural and anthropogenic pressures (Kim et al., 2013; Tsai et al., 2017). In this situation, supervised machine learning approaches, which make use of techniques from mathematical programming and statistics, have been used to scrutinize and model the environmental requirements of relevant macroinvertebrate taxa, and these techniques include decision trees (C4.5 – D'heygere et al., 2003) or multilayer perceptrons (Edia et al., 2010).

In addition, macroinvertebrate datasets include numerous taxa and a large number of samples, which can also cause difficulties for community analysis and river regionalization (Kim et al., 2013). In particular, ordination techniques and unsupervised machine learning approaches have been used to explore patterns of occurrence and community shifts and their relationships with environmental factors (Adriaenssens et al., 2007; Giraudel and Lek, 2001; Zhang et al., 2012a). Nevertheless, each ordination technique may have important limitations and assumptions that are incompatible with the over-dispersion and nonlinear nature of ecological data (Paliy and Shankar, 2016). Researchers have advocated for the use of a type of unsupervised artificial neural network called Kohonen Self-Organizing Maps (SOMs) (Kohonen, 1982), which have been demonstrated to be particularly competent in analyses such as macroinvertebrate community delineations (Chon, 2011; Park et al., 2007; Kim et al., 2013; Sroczyńska et al., 2017).

The freshwater ecosystems of China are a clear example of the abovementioned human-induced impacts. For instance, >40% of the rivers in China are notably polluted, which has led to poor drinking water quality for approximately 300 million rural residents (Liu and Yang, 2012). The river ecosystems in the northeast have also degraded due to industrial and agricultural development; thus, some river restoration work has been conducted in this area (Kong et al., 2013; Zhang et al., 2011; Zhang et al., 2013). The Hun-Tai River Basin is a large river basin with a basin area of 2.73×10^4 km^2. It represents the overall status of the water in the Liaohe River Basin in northeast China and is undergoing degradation. Many field surveys and studies using macroinvertebrates as important indicators in river health assessments showed the ecosystem were not in good conditions (Qu et al., 2016; Zhang et al., 2011; Zhang et al., 2013). However, most of these studies have mainly focused on small rivers or tributaries, and cannot reflect the overall status of the whole watershed, especially in such a large river basin. Up to now, little is known about the macroinvertebrate community and related environmental variables in the entire basin. We hypothesized when data from a large river basin and temporal span is merged, the geographical features could override the local environmental variables (*e.g.* water quality) in structuring the macroinvertebrate community, because the geographical features (*e.g.* elevation, distance from the river source) usually determine the macroinvertebrate structures when communities are studied at broader scales (Dedieu et al., 2014; Gaston, 2000).

This study analyzed the macroinvertebrate assemblages present in the large Hun-Tai River Basin to elucidate the existence of different types of communities and determine the indicator families and main environmental predictors for their occurrence. SOMs were used to reveal the existence of these macroinvertebrate communities (*i.e.*, clusters) in different areas of the river basin. Then, a genetically optimized C5.0 algorithm (Quinlan, 1992) (*i.e.*, a type of decision tree) was used to reveal the most important set of environmental predictors and indicator taxa of each community in the Hun-Tai River Basin.

2. Materials and methods

2.1. Study area

The Hun-Tai River is located in Liaoning Province of Northeast China, and it has two main tributaries, the Taizi and the Hunhe Rivers (Fig. 1). The lengths of the Hunhe and Taizi Rivers are approximately 415 and 413 km, respectively. The climate in this area is typical continental monsoon, with the highest temperature (34.3 °C) in the summer and lowest temperature (−25.2 °C) in the winter. The precipitation follows the temperature pattern, with the annual average precipitation 778 mm, 63% of which occurs in summer (Bu et al., 2014).

Field surveys were carried out in May 2009 (spring), August 2009 (summer), October 2009 (autumn) and May 2010 (spring). These surveys encompassed the entire river basin. In total, 287 sites from May 2009 to May 2010 in the Hun-Tai River Basin, in which 68 of Taizi River Basin were sampled twice (The number of sampling sites in each river and the codes of each river are located can be found in Appendix A). Consequently, 355 samples were ultimately collected where the selected environmental variables were measured *in situ* or obtained from reference databases.

2.2. Data collection

2.2.1. Environmental variables

All sampling sites were characterized using variables determining the geography, hydrology, climate, landuse, water quality and habitat, and there were 30 variables and 1 binary control variable (wadeable or non-wadeable) in total (Table 1). For each site, a handheld global positioning system (GPS, Trimble Juno SA) was used to obtain the latitude, altitude and elevation above sea level (m a.s.l.). Distance from the source was extracted from the digital map of the river basin using