Science of the Total Environment xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Science of the Total Environment

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

A data-mining framework for exploring the multi-relation between fish species and water quality through self-organizing map

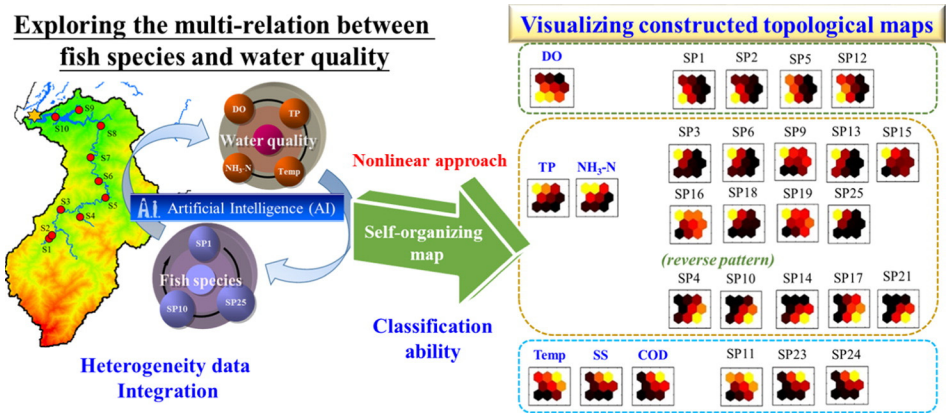
Wen-Ping Tsai [a], Shih-Pin Huang [b], Su-Ting Cheng [a], Kwang-Tsao Shao [b], Fi-John Chang [a,*]

[a] Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei 10617, Taiwan, ROC
[b] Biodiversity Research Center, Academia Sinica, Taiwan, ROC

HIGHLIGHTS

- Explore the relationship between fish community and water quality by AI techniques.
- SOM suitably reflects the spatial characteristics of fishery sampling sites.
- SOM visually classifies sampling data into 3 meaningful eco-water quality groups.
- Offer a cost-effective way to identify key water quality parameters and fish species.
- Provide a guiding reference relating fish species and water quality.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 12 September 2016
Received in revised form 17 October 2016
Accepted 11 November 2016
Available online xxxx

Editor: D. Barcelo

Keywords:
Eco-hydrological environments
Water quality
Flow regime
Fish community
Artificial neural network (ANN)
Self-organizing map (SOM)

ABSTRACT

The steep slopes of rivers can easily lead to large variations in river water quality during typhoon seasons in Taiwan, which may poses significant impacts on riverine eco-hydrological environments. This study aims to investigate the relationship between fish communities and water quality by using artificial neural networks (ANNs) for comprehending the upstream eco-hydrological system in northern Taiwan. We collected a total of 276 heterogeneous datasets with 8 water quality parameters and 25 fish species from 10 sampling sites. The self-organizing feature map (SOM) was used to cluster, analyze and visualize the heterogeneous datasets. Furthermore, the structuring index (SI) was adopted to determine the relative importance of each input variable of the SOM and identify the indicator factors. The clustering results showed that the SOM could suitably reflect the spatial characteristics of fishery sampling sites. Besides, the patterns of water quality parameters and fish species could be distinguishably (visually) classified into three eco-water quality groups: 1) typical upstream freshwater fishes that depended the most on dissolved oxygen (DO); 2) typical middle-lower reach riverine freshwater fishes that depended the most on total phosphorus (TP) and ammonia nitrogen; and 3) low lands or pond (reservoirs) freshwater fishes that depended the most on water temperature, suspended solids and chemical oxygen demand. According to the results of the SI, the representative indicators of water quality parameters and fish species consisted of DO, TP and *Onychostoma barbatulum*. This grouping result suggested that the methodology can be used as a guiding reference to comprehensively relate ecology to water quality. Our methods offer a cost-effective alternative to more traditional methods for identifying key water quality factors relating to

* Corresponding author.
E-mail address: changfj@ntu.edu.tw (F.-J. Chang).

http://dx.doi.org/10.1016/j.scitotenv.2016.11.071

fish species. In addition, visualizing the constructed topological maps of the SOM could produce detailed interrelation between water quality and the fish species of stream habitat units.

1. Introduction

Water resources management is closely tied up with hydrological and environmental conditions. River flow regimes, water quality and fish communities are crucial factors shaping river ecosystems, which need to be considered for sustainable management of water resources (Chang et al., 2009, 2011, 2013; Cheng et al., 2012; Kao et al., 2013; Lorenz et al., 1997; Pedro-Monzonís et al., 2016; Sun et al., 2016; Suen and Herricks, 2009; Tsai et al., 2015). The emphasis on water quality was driven initially by the societal concern about water pollution (Bennett et al., 2011; Campana et al., 2014). Assessments guidelines for water quality including physical, chemical and biological elements have been established to ensure a comprehensive water quality standard suitable not only for human beings but also for the whole inhabitant aquatic biomes in recent decades (Assessment, 2005; Cabral et al., 2012; Carr and Neary, 2008; Chapman, 1996). There are various evaluation indexes for assessing river water quality (Abdul Hameed M Jawad et al., 2010; Khanna et al., 2013). In Taiwan, Chen (2009) first divided the Danshui River into upstream, midstream and downstream areas based on geographic locations and then analyzed 16 water quality parameters in the river through statistical analyses with respect to the River Pollution Index (RPI) and the Water Quality Index (WQI).

Many studies linked the changes of fish communities with different environmental problems (Chang et al., 2011, 2013; Piperac et al., 2016; Segurado et al., 2016; Xu et al., 2016), such as excessive wastes due to expanding industrial as well as agricultural development and increasing population, which threatened inhabitant fish communities (Cronin et al., 2002; Hubert and Renno, 2006; Liu et al., 2012). Global warming as well as human activities have caused severe water deficiency and polluted watersheds, making management of water resources towards sustainability even more challenging nowadays. Decision makers need scientific-based information to maintain ecosystem services with a thorough consideration of ecosystem processes including changes in nature and alterations made by human activities (Assessment, 2005; Bateman et al., 2011). In Taiwan, freshwater fishes are of high diversity. At present, 10 freshwater fish species are listed as protected species while 42 freshwater fish species are listed as vulnerable or threatened species in consideration of habitat destruction, water pollution and violent competition with invasive fish species (Chen et al., 2012).

Despite the well-understood influence of flow regime on river ecosystems, rare studies focused on assessing the impacts of water quality and fish communities on river systems and their interactive relationships due to data scarcity as well as challenges originated from heterogeneous datasets. However, changes in water quality and fish communities are of equal importance, if not greater than, as flow regimes cause changes in ecosystems. This is especially true when minor changes of water quality can actually cause major shifts in fish communities and the associated ecosystem goods and services (Cosgrove and Loucks, 2015; Ghilarov, 2000; Groom et al., 2006). Most studies used statistical approaches to find correlations between water quality parameters and specific fish species (e.g. Cade and Noon, 2003; Kail et al., 2012; Kwon et al., 2012; Moerke and Lamberti, 2006). Nonetheless, studies exploring the interactive multi-relationships among various water quality parameters and fish communities, to our best knowledge, are very limited. As a result, exploring complex relationships between water quality parameters and fish species communities requires the development of innovative techniques.

The self-organizing map (SOM) is one of the powerful artificial intelligence (AI) techniques that suitably relate similar features of large dimensional dataset to a reduced dimension set of output responses in topological orders (Chang et al., 2013; Tsai et al., 2016; Hsu et al., 2002; Kalteh et al., 2008; Liao et al., 2005; Liu et al., 2006; Vesanto, 1999; Vesanto et al., 2000). Alvarez-Guerra et al. (2008) stated that the SOM could be an effective tool to classify datasets according to their similar sediment characteristics coupled with the integration of multiple physical, chemical and ecotoxicological variables. Grieu et al. (2006) used the SOM neural network to analyze and estimate chemical parameters. With growing environmental awareness, identifying the relationship between water quality and fish species has gained considerable attention, and thus has become an important issue for achieving a full consideration of the environmental and ecological changes associated with ecosystem managements and decisions (Keeler et al., 2012). So far, the SOM has rarely been used to explore the relationship between water quality factors and fish species.

In this study, we aimed to extract the complex multi-relation between water quality factors and fish community and provide visible evidence regarding eco-systems. We first explored the principle interrelations by employing the SOM based on long-term heterogeneous datasets. We then identified the indicators of the environmental condition from various water quality parameters and fish species. Consequently, a holistic ecosystem management incorporating water quality and fish species factors for the conservation on threatened freshwater fishes in riverine could be developed.

2. Study area & data description

2.1. Study area

The Danshui River, the longest river in northern Taiwan, has a total length of 159 km and occupies a drainage area of 2726 km^2. The Shihmen Reservoir, built in 1964, is the pivotal hydraulic facility in the Danshui River basin for irrigation, domestic and industry uses as well as hydropower generation and flood control purposes. This study chose the Dahan River (i.e. the Shihmen Reservoir watershed) as the study river reach (Fig. 1). The Dahan River is the main upstream tributary of the Danshui River, and its length and catchment area are of 135 km and 1163 km^2, respectively. The Dahan River originates from the central mountains at an elevation more than 3000 m and then flows for a distance of 135 km into the Danshui River at an elevation of 100 m. The average streamflow in the Dahan River is 163 cms, and the average annual rainfall in the Shihmen Reservoir watershed is 2410 mm (over 70% occurs in typhoon seasons during May and October).

2.2. Data description

In this study, most fish sampling accompanied with water quality survey was conducted seasonally, mainly in February, May, August and October during 2003 and 2013 (except 2009 due to no investigative data), by the North Water Resources Bureau in Taiwan. However, sampling frequencies were not consistent among sites due to funding and labor constraints. For example, sites like S2 and S3 were not sampled as frequent as other sites. Also, in year 2009, there were unfortunately no investigations at all. Thus, there were a total of 276 seasonal datasets with 33 input variable (25 fish species (SP1–SP25) and 8 water quality parameters, Table 1) collected at 10 sampling sites along the Dahan River. The sampling sites are shown in Fig. 1.

These eight water quality parameters consisted of water temperature (Temp), hydrogen ions concentration (pH), electric conductivity (EC), dissolved oxygen (DO), suspended solids (SS), chemical oxygen

Please cite this article as: Tsai, W.-P., et al., A data-mining framework for exploring the multi-relation between fish species and water quality through self-organizing map, Sci Total Environ (2016), http://dx.doi.org/10.1016/j.scitotenv.2016.11.071