Journal of Hydrology 420–421 (2012) 292–300

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

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

Recognition of key regions for restoration of phytoplankton communities in the Huai River basin, China

Changsen Zhao [a,b,*], Changming Liu [a], Jun Xia [a], Yongyong Zhang [a], Qiang Yu [b], Derek Eamus [c]

[a] Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, PR China

[b] Plant Functional Biology and Climate Change Cluster, University of Technology, Sydney, Broadway, NSW 2007, Australia

[c] National Centre for Groundwater Research and Training, University of Technology Sydney, Broadway, NSW 2007, Australia

ARTICLE INFO

Article history:
Received 25 July 2011
Received in revised form 6 December 2011
Accepted 7 December 2011
Available online 17 December 2011
This manuscript was handled by Geoff Syme, Editor-in-Chief

Keywords:
Key region recognition
Random forests clustering
Biodiversity
Ecological niche
Phytoplankton
Water chemistry

SUMMARY

Healthy phytoplankton communities are the basis of healthy water ecosystems, and form the foundation of many freshwater food webs. Globally many freshwater ecosystems are degraded because of intensive human activities, so water ecosystem restoration is a burning issue worldwide. Selection of key regions for phytoplankton-related restoration is crucial for an effective aquatic eco-restoration. This paper presents a practical method for identification of key regions for phytoplankton-related restoration, using random forests (RFs) method to cluster sites based on dominance, biodiversity, water chemistry and ecological niche. We sampled phytoplankton for species richness and relative abundance and water quality in the Huai River basin (HRB), China to determine the phytoplankton communities' composition and structure and characterize of their ecological niches. A wider mean niche breadth of a species usually leads to a greater overlap with the niche of other species. Using these data and water quality indices, we identified the key regions for phytoplankton-related river restoration activities. Results indicate that our method for recognition of key regions is effective and practical and its application to the HRB identified the Northern Plain area as the key region for restoration. This area is severely polluted and contributes significantly to the HRB phytoplankton communities. Phytoplankton in this region is highly adaptable to environmental change and therefore will be relatively unharmed by environmental instability induced by restoration measures. During restoration, indices of water temperature, total phosphorus and chemical oxygen demand can be altered with little negative influence on phytoplankton communities, but measures that increase ammonia–nitrogen concentration would be highly detrimental. These results will provide valuable information for policy makers and stakeholders in water ecosystem restoration and sustainable basin management in the HRB.

1. Introduction

The world's ecosystems (both terrestrial and aquatic) are capital assets. They can yield numerous vital services, including the production of goods, life support processes, and life-fulfilling conditions when properly managed. Ecosystem services are essential to human existence and operate on such a grand scale, and in such intricate and little-explored ways, that most could not be replaced by technology. Unfortunately, escalating impacts of human activities on ecosystems imperil their delivery. Ecosystems are undergoing rapid degradation and depletion (Daily, 1999; Daily et al., 2000) and are increasingly threatened by human-induced habitat loss (Kagalou et al., 2010). With economic development and population increase over the past decades, large quantities of pollutants have been discharged into rivers in China. This has resulted in degradation of water quality and greatly impaired aquatic ecosystems, which, in turn, has severely hindered the sustainable development of the economy and society. This is especially true for the Huai River, China.

The Huai River, located between the Yangtze River and the Yellow River, forms a geographical divide between northern and southern of China. The Huai River basin (HRB) is one of the main grain-producing areas of China. Its population (0.165 billion by the year 2000) surpasses all other large basins in China (Zhao et al., 2010). Over the last half century, it has been severely affected by human activities, especially construction of dams and weirs and discharge of pollutants. Because of serious flood disasters and flood control requirements in the Huai River, around 11,000 dams and sluices had been built by the year 2000. The number of such structures on this river accounts for approximately half of those in China and a quarter of those in the world (Liu et al., 2011). These structures have brought tremendous economic benefit through flood control, increased irrigation, and power

* Corresponding author at: Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, PR China. Tel./fax: +86 10 64889083.
E-mail address: hzjohnson@gmail.com (C. Zhao).

0022-1694/$ - see front matter Crown Copyright © 2011 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2011.12.016

generation in the basin. However, major counterarguments have been raised for many years regarding their detrimental impacts on the environment (Wang and Xia, 2010; Zhao et al., 2010). Dams and sluices usually lead to hydrologic "fragmentation", which greatly harms aquatic ecosystems of the HRB. Water quality in more than 83% of rivers cannot reach the national criteria (GB3838-2002) and the quality in this basin is the worst in the nation's top seven basins (based on Chinese Environment Bulletin in 2005) (Zhang et al., 2010). The river has a history of disastrous pollution events. One severe water pollution event happened in 1994 because of major floods, made worse by inappropriate operation of sluices. Large number of fish and other aquatic organisms were killed, and there was serious damage to ecosystems in regional rivers. Affected waterworks had to stop supplying water for 54 days and 1.5 million people suffered from a shortage of drinking water. This caused economic losses of at least US$200 million (Xia et al., 2011). Long-term mismanagement of dams and sluices for water use, plus excessive pollution discharge, has resulted in ecosystems in many middle and lower river reaches that are seriously degraded and extremely unstable (Zhao et al., 2008). Consequently there is a need to re-construct or restore stable, healthy aquatic ecosystems. The top priority should be given to the organisms that underpin the food webs of aquatic ecosystems – phytoplankton communities in the HRB.

In aquatic ecosystems, phytoplankton is the major microbial biomass. Light energy conversion and related synthesis of carbon compounds is carried out by three major primary producers – higher plants (macrophytes), phytoplankton, and photosynthetic bacteria. Phytoplankton are the main microorganisms involved in this process (Sigee, 2005). They play a major role in aquatic ecosystems as their biological activity affects the biogeochemical cycles of a number of macro and micronutrients (carbon, silicon, sulfur, nitrogen, iron, etc.) (Falkowski, 1994; Falkowski et al., 1998). Phytoplankton are also the principal primary producer of freshwater food webs and mainly depends on light energy and nutrients. However, too many nutrients often accelerate the growth of phytoplankton and an overpopulation of phytoplankton is extremely harmful to the local aquatic-related economy. For example, toxic algal blooms worsen water quality and greatly impair local fisheries. The abundance and biomass of phytoplankton should therefore be controlled within a moderate range. An index of "dominance" (Zhao et al., 2011), which denotes the importance or contribution of a species can be used to assess community structure and used to manage aquatic ecosystem health. To control the "dominance" of a species, understanding of the connection between its abundance and its habitat indices, especially indices of water quality are necessary.

The ecological niche can establish connections between phytoplankton species and indices of water pollution. It is one of the most important concepts in the exploration of biological communities' structure and development, biodiversity, association of species with a particular environment, conservation planning and decision making (Pearce and Lindenmayer, 1998; Ferrier, 2002; Wiley et al., 2003). Hutchinson (1957) formalized the niche as an n-dimensional hyper volume whose axes are critical physical and environmental factors determining the existence of a species – this concept of niche as a function of measurable factors has provided a foundation for many theoretical and field studies (Smith, 1982).

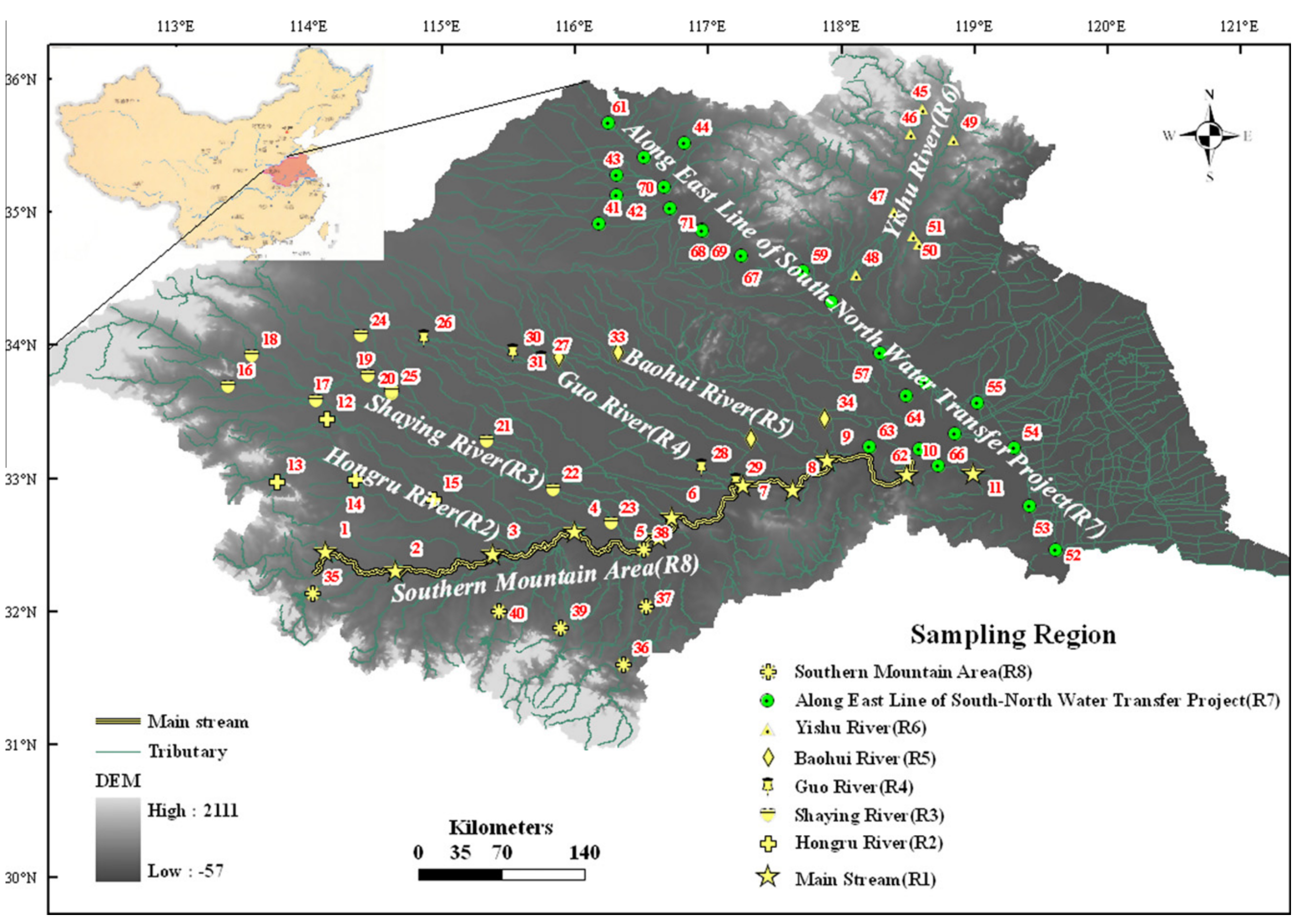

Fig. 1. Phytoplankton sampling in the HRB (modified from Zhao et al. (2011)), the numbers on the map indicate the sampling sites.

Table 1
Selected water chemistry indices in the Huai River basin.

No.	Environmental factor	Range	Mean ± SD	Mode
1	Water temperature (WT)	27.3–33.4	29.9 ± 1.34	30.6(8.82%)
2	pH	6.08–8.09	7.33 ± 0.39	7.55(5.63%)
3	Dissolved oxygen (DO)	1.20–11.8	5.76 ± 2.37	7.80(11.1%)
4	Ammonia nitrogen (NH_3–N)	0.14–12.4	1.13 ± 2.39	0.27(7.04%)
5	Permanganate index (COD_{Mn})	1.96–18.7	5.58 ± 2.78	4.40(7.04%)
6	Chemical oxygen demand (COD_{Cr})	11.1–54.7	22.6 ± 10.1	13.5(6.45%)
7	Total phosphorus (TP)	0.017–1.742	0.16 ± 0.26	0.03(8.45%)
8	Total nitrogen (TN)	0.32–15.4	2.66 ± 2.97	3.07(4.23%)
9	Fluoride	0.5–0.9	0.65 ± 0.17	0.50(52.1%)

Unit of water temperature (WT) is degrees Celsius, pH has no unit and the rest are all in mg L^{-1}.
The mode value in the table is expressed as "Mode (percentage of mode number to total number)".

Studies on ecological niche have mainly focused on niche breadth and niche overlap (Thompson and Gaston, 1999; Jehle et al., 2000; Brändle et al., 2002). Ecological niche has been widely applied in research on habitat selection, species conservation, spatial distribution and temporal dynamics, temporal and spatial niche-partitioning, species delimitation, exotic species invasion and community succession, etc. (McNyset, 2005; Domínguez-Domínguez et al., 2006; Chen et al., 2007; Irfan-Ullah et al., 2007; Peterson et al., 2007; Raxworthy et al., 2007; Solano and Feria, 2007; Basille et al., 2008; Foulon et al., 2008; Friberg et al., 2008; Peterson and Nakazawa, 2008; Quero et al., 2008; Thorn et al., 2009; Waltari and Guralnick, 2009).

The protection or restoration of whole ecosystems often represents the most effective way to sustain genetic, population, and species diversity (Vitousek et al., 1997). Understanding the consequences of biodiversity changes on ecosystem functioning is becoming increasingly critical in view of the profound influence of human activity on natural ecosystems and the goods and services humans receive from them (Vitousek et al., 1997; Daily et al., 2000; Giller et al., 2004). To achieve restoration of aquatic ecosystems across a very large area, selection of appropriate key regions or priority areas is essential.

Most studies in aquatic systems have used plants, invertebrates, fish and birds as indicators (Altaba, 1990; Crandall, 1998; Posadas and Crisci, 2001; Pérez-Losada et al., 2002; Turpie et al., 2002; Sánchez-Fernández et al., 2004; Abellán et al., 2005). Few studies focused on the phytoplankton, and furthermore, few considered many factors of dominance, biodiversity, water quality and ecological niche.

In view of the major role of phytoplankton in aquatic ecosystems, the objective of this paper is to identify the key regions for future restoration of degenerated phytoplankton communities resulting from severe water pollution. This study was based on assessments of dominance, biodiversity and ecological niche along a gradient of water quality indices.

2. Methodology

2.1. The study area

The Huai River is the sixth largest river in China. It is located between the Yangtze River and the Yellow River of China (Wang and Xia, 2010). The area of the Huai River basin (HRB: 30°55′–36°36′N, 111°55′–121°25′E) covers 27,000 km^2. The HRB lies at China's transition between the northern climate and southern climate (Gao et al., 2010). It is the most densely inhabited river basin and the main grain-producing area of China. In 2005, the total population and grain yield accounted for 13.1% and 16.1%, respectively, of the national total (Xia et al., 2011). The population density surpasses all other large basins in China (Zhao et al., 2010).

The HRB can be divided into eight regions: Main Stream (R1), Hongru River (R2), Shaying River (R3), Guo River (R4), Baohui River (R5), Yishu River (R6), Along East line of South-North Water Transfer Project (R7), Southern Mountain Area (R8), as shown in Fig. 1.

2.2. Methods

To explore the impacts of water pollution on phytoplankton communities and therefore identify key regions for ecological restoration in the HRB, we sampled phytoplankton as well as water chemistry at 71 typical sites (Fig. 1). We sampled every day during a low-water period in the hot wet season (from July 10th to July 20th, 2008). This low-water period is the best for exploration of relationships between phytoplankton and water quality because during this period phytoplankton flourish and the impacts of dam on water quality and water ecosystems are greatly reduced since most water sluices are kept open.

2.2.1. Water chemistry

We measured *in situ* indices of water chemistry indices: water temperature: WT; pH; and dissolved oxygen: DO by using a portable HACH PC101. Twenty-one additional indices were tested in the laboratory. Water samples were sent to laboratory within 24 h. Various instruments were used to analyze the additional indices. Among these instruments, Spectrophotometer (DR5000) was used to measure Ammonia nitrogen (NH_3–N), total phosphorus (TP), total nitrogen (TN) and hexavalent chromium; Atomic Absorption Spectrophotometer (Thermo M6) was used for test of copper (Cu), zinc (Zn), cadmium (Cd) and lead (Pb); Ion Chromatograph (DIONEX-600) was employed to measure sulfate, fluoride, chloride and nitrate; Automatic Flow Injection Analyzer (KALAR SAN^{++}) to measure cyanide, volatile phenol, anionic detergent. Of the initial 21 water chemistry indices selected for analyses the concentrations of many of them were at or below the limits of detection at 69.0–100% of the sampling sites. Consequently we selected nine main factors (WT; pH; DO; NH_3–N; permanganate index: COD_{Mn}; chemical oxygen demand: COD_{Cr}; TP; TN; Fluoride) as the water chemistry indicators of the Huai River (Table 1).

2.2.2. Phytoplankton sampling and taxa determination

A 1000 mL-capacity organic glass bottle was used to sample water from 0 to 2 m below the water surface. As quickly as possible, 1.5% concentration Lugol's solution was added to the bottle. In the laboratory, a 24-h sedimentation method was used to concentrate the phytoplankton sample to 30 mL. A 0.1 mL sub-sample was taken from the 30-mL concentrated sample and loaded into a 0.1-mL plankton counting chamber. Finally, the phytoplankton were counted using Utermöhl's inverted plankton microscope. The biomass was converted from biovolume assuming a specific gravity of 1.0. To determine individual biovolume, individual size (length, height and breadth, or diameter) of a species was measured with the plankton microscope. Average size of at least 50 individuals was used to calculate average biovolume of a species (SL167-96[1]).

2.2.3. Dominance assessment

Abundance and biomass are fundamental indices for biological monitoring. The two indices often rank differently, which makes it hard to objectively assess the dominance or importance of a species in a community (Zhao et al., 2011). To overcome this, Zhao et al. (2011) combined them into one index by using the following equation:

$$I_{importance} = \omega_1 PCT_{abundance} + \omega_2 PCT_{biomass} \tag{1}$$

where $I_{importance}$ stands for the dominance of a species; $PCT_{abundance}$ and $PCT_{biomass}$ refer to the ratio of the species' abundance and biomass to the total for the communities, respectively; ω_1 and ω_2 are the weightings of abundance and biomass, and let $\omega_1 = \omega_2 = 0.5$.

The larger the $I_{importance}$ is, the more the species contributes to its community, and the more important it is in the community.

2.2.4. Biodiversity

We employed the commonly used Shannon Index (H) (Spellerberg and Fedor, 2003; Shannon and Weaver, 1949):

$$H = -\sum_{i=1}^{s} \left[\left(\frac{n_i}{N} \right) \ln \left(\frac{n_i}{N} \right) \right] \tag{2}$$

where H stands for the biodiversity; n_i refers to the number the ith species, in [individual L^{-1}]; N is the total number of all species in a sample, in [individual L^{-1}]; s refers to the species type number in a sample. When all species are equally abundant, H reaches its peak value.

2.2.5. Niche breadth and niche overlap

There are many models to calculate niche breadth and overlap (Levins, 1968; Pianka, 1974; Hurlbert, 1978; Smith, 1982). In this paper, we employ the widely-used Levins Breadth Model (Levins, 1968; Eq. (3)) and Pianka Overlap Model (Pianka, 1974; Eq. (4)) to get niche breadth and niche overlap, respectively.

$$\text{Levins}' \text{ Breadth Model}: B_i = 1 \Big/ \sum_{j=1}^{R} (P_{ij})^2 \tag{3}$$

where B_i is the niche breadth of species i; P_{ij} stands for the ratio of the number of individuals of species i in resource state j to the total number of individuals of species i. R refers to the total number of resource states. Resource states are defined according to national water quality criteria. They stand for gradients along one available resource. Resources available includes biochemical oxygen demand (BOD5), dissolved oxygen (DO), permanganate index (COD_{Mn}), ammoniacal nitrogen (NH_3–N), and so on.

$$\text{Pianka Overlap Model}: \; O_{ik} = \sum_{j=1}^{R} P_{ij} P_{kj} \Bigg/ \sqrt{\sum_{j=1}^{R} P_{ij}^2 \sum_{j=1}^{R} P_{kj}^2} \tag{4}$$

where O_{ik} is the niche overlap of species i on species k; P_{ij} and P_{kj} are respectively the ratios of numbers of individuals of species i and species k in resource state j to the total number of individuals of species i and k; $O_{ik} \neq O_{ki}$. Calculation of niche breadth and niche overlap was conducted using the software "Data Processing System (DPS)" (Tang and Feng, 2007).

2.2.6. Random forests (RFs) and statistical methods

There are many methods available for classification of sampling sites. Dudoit et al. (2002) compared and reviewed these (Mehrian et al., 2007). Clustering is important for pattern recognition, classification, model reduction and optimization (Hardin and Rocke, 2004; Shafi et al., 2010). Most clustering algorithms require as input a dissimilarity measure between samples (Shi et al., 2005) whereas an unsupervised learning method – random forests (RFs) – does not require this. An RF predictor is an ensemble of individual classification tree predictors (Breiman, 2001). For each observation, each individual tree falls into one class and the forest predicts the class that has the largest number of trees. The user has to specify the number of randomly selected variables *mtry* to be searched through for the best split at each node (Horvath et al., 2007). Injecting the right kind of randomness makes RFs accurate classifiers. Using out-of-bag estimation makes concrete the otherwise theoretical values of strength and correlation. Random inputs and random features produce good results in classification (Breiman, 2001).

In addition, RFs have a number of theoretical advantages (Shi et al., 2005). First, the clustering results do not change when one or more covariates are monotonically transformed since the dissimilarity only depends on the feature ranks and one does not need to worry about symmetrizing skewed covariate distributions. Second, the random forest dissimilarity weighs the contributions of each covariate on the dissimilarity in a natural way: the more related the covariate is to other covariates, the more it will affect the definition of the random forest dissimilarity. Third, the random forest dissimilarity does not require the user to specify threshold values for dichotomizing expressions. It automatically dichotomizes the expressions in a consistent, data-driven way based on individual tree predictors. Fourth, the random forest dissimilarity naturally accommodates missing values.

The research of Horvath et al. (2007) indicates that the results of a RF analysis are highly robust with respect to the RF parameter *mtry* (the number of variables considered at each split). The default value for the number of random features is the square root of the number of variables. A low value of *mtry* is appropriate when most variables are highly correlated to the outcome. If there is no a *priori* knowledge, to choose a high value of *mtry* is a better choice. Also, Horvath et al. (2007) found that large values of *mtry* (5000) leads to higher prediction accuracy and a lot of trees (30,000) make the estimate of the importance measure stable.

In this paper, RFs were employed to cluster the 71 sampling sites, taking as inputs a total of 29 variables (dominance, biodiversity, nine water quality indices, nine niche breadth and nine niche overlap variables along the water chemistry indices' gradients). The sites were regarded as points in two-dimensional multidimensional scaling plots. Then the characteristics of every cluster were identified, based on which key regions were recognized.

To test whether the differences between clusters were significant, a nonparametric test, the Kruskal–Wallis Test, was employed in our study. This test neither makes assumptions for a distribution, nor assumes that any particular distribution is being used. It is often used in tests for multiple independent samples with different sample sizes. Every RF clustering was accompanied by a Kruskal–Wallis Test. The clustering result with minimum asymptotic significance (P-value) was selected as the optimized outcome.

The RFs clustering and Kruskal–Wallis Test analyses were conducted with the freely available software R (R Development Core Team, 2011).

[1] "Standard for the investigation of reservoir fishery resources" issued by the Ministry of Water Resources of China.

3. Results and discussion

3.1. Phytoplankton community structure and representative species selection

There were, in total, 5 classes and 39 species in the sampled Huai River phytoplankton communities. Among them were 17 species of *Chlorophyta*; 12 *Cyanophyta*; 6 *Bacillariophyta*; 2 *Euglenophyta*; and 2 *Pyrrophyta*. *Cyanophyta* was the overwhelmingly dominant class in the HRB phytoplankton communities (94.38% of total abundance); *Chlorophyta* ranked second but only accounted for 5.38% of the total abundance; *Euglenophyta* for 3%, *Pyrrophyta*, for 2% and *Bacillariophyta*, for 0.19%.

Among the *Cyanophyta*, *Aphanizomenon flos-aquae* had the greatest abundance (accounting for 40.00% of the total in the HRB phytoplankton communities) while *Oscillatoria* had the greatest biomass (61.32% of total communities). On the whole, *A. flos-aquae* and *Oscillatoria* made the largest contribution in terms of their dominance values in the HRB phytoplankton communities (24.27% and 39.01%, respectively). The 10 species with dominance values larger than 1% are listed in Table 2. Their dominance values summed to 92.71%, so they were selected to represent the whole HRB phytoplankton communities in the study period.

3.2. Ecological niche of the representative species

We computed the ecological niche breadth of the 10 representative species along the nine main water chemistry indices (Table 3) based on Eq. (3).

Generally, a species with a wider niche breadth has a greater adaptability while one with a narrower niche breadth is sensitive to environmental change. The former often has a much greater chance of survival than the latter under conditions of limited resources; however, the latter is usually more competitive in its local habitat when resources are abundant because of its higher efficiency in use of resources (Chen et al., 2009).

Among the 10 species, *Oscillatoria* had the broadest mean breadth value (3.536), suggesting greater adaptability in a changing environment. It had the broadest niche breadth along the TP gradient (5.252) and the narrowest one along TN (2.341). In contrast, *Anabaena azotica* had the narrowest mean niche breadth (1.889), signifying a poor adaptability to changing environmental conditions.

The 10 species had broader niche breadths along the WT, TP and COD_{Cr} gradients on average (means: 4.143, 3.465 and 3.368, respectively). The narrowest breadth on average was along the NH_3–N gradient (mean: 1.965). Data analysis indicated that most species in the study area favored a habitat with a lower NH_3–N concentration in the range 0.15–0.5 mg L^{-1}. This means that changes of WT, TP and COD_{Cr} will have few negative impacts on the phytoplankton communities in the HRB; while any small increase in NH_3–N concentration might cause great instability of the phytoplankton communities.

These results imply that during restoration of phytoplankton communities in the HRB, water temperature, concentration of TP and COD_{Cr} can be altered, but measures that would increase NH_3–N concentration should be avoided.

Generally, if two species have a large niche overlap, this suggests that they have a similar behavior in utilization of resources, which might result in strong competition with each others under some conditions.

Phytoplankton species in the HRB had the largest niche overlap with each other along the NH_3–N axis (mean: 9.034 while breadth along NH_3–N was ranked last. This means the representative species had the narrowest niche breadth along it but all were concentrated in a narrow NH_3–N gradient, as discussed above. Compared with niche breadth (Table 3), niche overlap in Table 4 approximates a reverse trend, which implies that species in the HRB are coexisting with each other with little competition for WT, TP, COD_{Cr} and DO.

The species with the lowest dominance, *Dactylococcopsis acicularis*, had the largest overlap value along the NH_3–N gradient (9.4297) and the least overlap along the DO gradient (8.3922). On the whole, the largest overlap values for most species occurred along the NH_3–N gradient, while more than half of the smallest overlaps were along the DO gradient.

The overlap values for each species are listed in Table 5. *D. acicularis* had the largest total overlap with other species. *Spirulina major* had the second largest overlap with others, while *A. azotica*, with the smallest mean breadth, had the least total overlap with other species. According to the research of Jiang et al. (2009), we concluded that *D. acicularis* has the greatest similarity in resource use with the others while *A. azotica* has the least similarity. Overall, the total overlap and mean breadth of the phytoplankton species in the Huai River have the same trend ($Total\_overlap = 1.13 * Mean\_breadth + 3.90$, $R^2 = 0.72$, $F = 20.64 \gg F_{0.01} = 5.35$) – a wider mean niche breadth of a species usually leads to a greater overlap with other species.

3.3. Spatially clustering sites using random forests

To make the estimate of importance measure more stable and keep a higher prediction accuracy, we chose a large number of trees (*ntree*: 30 000) for each random forest fit and used a large number of random features (*mtry*: 5000) as recommended by Horvath et al. (2007). The optimized outcome ($P = 0.02 < 0.05$ with the Kruskal–Wallis Test) showed that the 71 sites were grouped into six clusters with different sizes of 20, 10, 14, 13, 5 and 9 sites, respectively. The *P*-value indicated that there were significant dissimilarities among the six clusters. The optimized clustering results are shown in Fig. 2.

To facilitate study of the characteristics of every cluster, we divided every factor of dominance, biodiversity, water chemistry, mean ecological niche breadth and mean ecological niche overlap into six different grades: highest, higher, middle higher (MH), middle lower (ML), lower and lowest. Then we analyzed the characteristics of the six clusters. We found the following:

- *Cluster 1* was associated with lower-grade pollution, consisting of a middle-lower NH_3–N concentration, lower concentration of COD_{Mn}, COD_{Cr}, TP, TN and fluoride; a lower dominance and a higher biodiversity; a middle lower mean breadth value and a middle-higher mean overlap value.
- *Cluster 2* had the lowest pollution including the lowest concentration of NH_3–N, COD_{Mn}, TP, TN and fluoride plus a middle lower COD_{Cr} concentration; the lowest dominance and a lower biodiversity; a lower mean breadth and a middle lower mean overlap.

Table 2
Dominance values of representative species in the HRB phytoplankton communities.

Species	Dominance value (%)
Oscillatoria	39.01
Aphanizomenon flos-aquae	24.27
Spirulina major	10.72
Anabaena azotica	6.83
Microcystis aeruginosa	2.75
Pinnularia nobilis	2.24
Merismopedia glauca	2.12
Pediastrum simplex	1.99
Anabaena flos-aquae	1.43
Dactylococcopsis acicularis	1.35

Table 3
Niche breadth of the 10 representative species along the gradients of the nine water chemistry indices.

Species	WT	pH	DO	NH_3–N	COD_{Mn}	COD_{Cr}	TP	TN	Fluoride	Mean	Rank
Oscillatoria	4.562	2.490	4.371	2.660	3.019	4.562	5.252	2.341	2.572	3.536	1
Aphanizomenon flos-aquae	4.890	3.009	2.572	1.338	2.846	4.083	3.081	3.730	2.568	3.124	4
Spirulina major	4.845	3.019	3.834	2.530	2.828	3.415	4.310	2.729	3.307	3.424	2
Anabaena azotica	2.000	2.000	2.000	2.000	1.000	2.000	2.000	2.000	2.000	1.889	10
Microcystis aeruginosa	3.769	1.815	1.815	1.000	2.333	3.267	1.815	2.333	2.333	2.276	9
Pinnularia nobilis	4.500	2.571	3.600	2.571	3.000	3.000	3.000	3.000	2.571	3.090	5
Merismopedia glauca	4.741	2.844	2.560	1.488	2.612	3.657	3.122	2.977	2.844	2.983	7
Pediastrum simplex	2.667	1.600	2.667	1.600	1.600	1.600	4.000	2.667	4.000	2.489	8
Anabaena flos-aquae	4.840	2.575	2.200	2.283	2.575	3.903	3.457	2.575	3.270	3.075	6
Dactylococcopsis acicularis	4.613	2.844	2.926	2.179	3.141	4.197	4.613	2.522	3.580	3.402	3
Mean	4.143	2.477	2.854	1.965	2.495	3.368	3.465	2.687	2.905		
Rank	1	8	5	9	7	3	2	6	4		

WT: water temperature; DO: dissolved oxygen; NH_3–N: ammoniacal nitrogen; COD_{Mn}: permanganate index; COD_{Cr}: chemical oxygen demand; TP: total phosphorus; TN: total nitrogen.
Unit of WT is degrees Celsius, pH has no unit and the rest are all in mg L^{-1}.

Table 4
Niche overlap of the 10 representative species along the gradients of the nine chemistry indices.

Species	WT	pH	DO	NH_3–N	COD_{Mn}	COD_{Cr}	TP	TN	Fluoride	Mean	Rank
Oscillatoria	8.5086	8.7613	8.4017	9.2196	8.9804	8.4786	7.9708	9.022	8.4231	8.641	4
Aphanizomenon flos-aquae	8.5302	9.1848	8.319	9.3395	8.1564	8.2687	8.249	8.965	8.7145	8.636	5
Spirulina major	8.3563	8.6672	8.6199	9.3662	9.0509	8.3542	8.6162	8.9929	8.4595	8.720	2
Anabaena azotica	5.2063	9.1596	4.4143	7.4603	6.9076	6.2625	5.5216	8.6585	7.1865	6.753	10
Microcystis aeruginosa	7.4062	6.7036	7.8619	9.2439	7.2509	7.9711	6.8582	6.4978	8.9249	7.635	8
Pinnularia nobilis	6.8815	9.0279	5.1008	8.5123	8.4117	5.7369	5.9614	8.7738	9.0724	7.498	9
Merismopedia glauca	8.4863	9.2665	8.2106	9.357	9.0232	8.4834	7.7827	8.7735	8.871	8.695	3
Pediastrum simplex	6.7342	8.1397	8.1791	8.9925	7.3231	5.8189	8.0262	8.7504	8.4786	7.827	7
Anabaena flos-aquae	8.1681	8.7179	7.9855	9.4174	8.7889	7.621	8.4288	9.2537	9.18	8.618	6
Dactylococcopsis acicularis	8.5941	9.2267	8.3922	9.4297	9.0877	8.6559	8.4913	9.1246	9.1693	8.908	1
Mean	7.687	8.686	7.549	9.034	8.298	7.565	7.591	8.681	8.648		
Rank	6	2	9	1	5	8	7	3	4		

WT: water temperature; DO: dissolved oxygen; NH_3–N: ammoniacal nitrogen; COD_{Mn}: permanganate index; COD_{Cr}: chemical oxygen demand; TP: total phosphorus; TN: total nitrogen.
Unit of WT is degrees Celsius, pH has no unit and the rest are all in mg L^{-1}.

Table 5
Mean niche overlaps between the representative species.

Species	*Oscillatoria*	*Aphanizomenon flos-aquae*	*Spirulina maior*	*Anabaena azotica*	*Microcystis aeruginosa*	*Pinnularia nobilis*	*Merismopedia glauca*	*Pediastrum simplex*	*Anabaena flos-aquae*	*Dactylococcopsis acicularis*	Total with others
Oscillatoria	1.000	0.866	0.960	0.677	0.748	0.844	0.900	0.820	0.894	0.932	7.641
Aphanizomenon flos-aquae	0.866	1.000	0.889	0.621	0.923	0.735	0.943	0.810	0.902	0.947	7.636
Spirulina major	0.960	0.889	1.000	0.682	0.753	0.810	0.926	0.846	0.898	0.955	7.720
Anabaena azotica	0.677	0.621	0.682	1.000	0.463	0.635	0.654	0.616	0.700	0.705	5.753
Microcystis aeruginosa	0.748	0.923	0.753	0.463	1.000	0.623	0.813	0.648	0.837	0.827	6.635
Pinnularia nobilis	0.844	0.735	0.810	0.635	0.623	1.000	0.743	0.609	0.723	0.773	6.498
Merismopedia glauca	0.900	0.943	0.926	0.654	0.813	0.743	1.000	0.825	0.922	0.969	7.695
Pediastrum simplex	0.820	0.810	0.846	0.616	0.648	0.609	0.825	1.000	0.798	0.856	6.827
Anabaena flos-aquae	0.894	0.902	0.898	0.700	0.837	0.723	0.922	0.798	1.000	0.943	7.618
Dactylococcopsis acicularis	0.932	0.947	0.955	0.705	0.827	0.773	0.969	0.856	0.943	1.000	7.908

- *Cluster 3* had a middle-lower pollution level, with middle-lower concentrations of TP, TN and fluoride, a lower NH_3–N concentration, and the middle concentration of COD_{Mn} and COD_{Cr}. It also had a higher dominance but the lowest biodiversity in addition to the lowest mean breadth and overlap.
- *Cluster 4* had a higher pollution level, with higher concentrations of NH_3–N, COD_{Mn}, COD_{Cr}, TP and TN, plus the highest fluoride concentration. Besides, it had a middle higher dominance and a middle lower biodiversity. Its mean breadth ranked the highest in the six clusters and its mean overlap was higher.

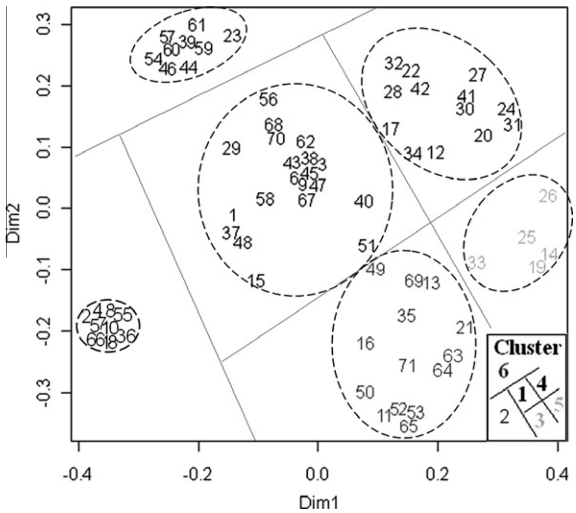

Fig. 2. Clusters by random forests based on dominance, biodiversity, water chemistry, ecological niche breadth and overlap. The numbers in the figure stand for the sampling site code, as in Fig. 1. Dim1 and Dim2 stand for the two dimensions of a point.

- *Cluster 5* had the highest pollution level because of the highest concentration of NH_3–N, COD_{Mn}, COD_{Cr}, TP and TN, and a higher fluoride concentration. It had the highest dominance and biodiversity. Its mean breadth was middle higher and its mean overlap was lower.
- *Cluster 6* had a middle higher pollution level, composed of middle higher concentrations of NH_3–N, TP, TN and fluoride, plus a middle lower COD_{Mn} and the lowest COD_{Cr} concentration. Additionally, it had a middle lower dominance and a middle higher biodiversity. Its mean breadth was higher and the mean overlap ranked the highest in the HRB phytoplankton communities during our study period.

The spatial distribution of the six clusters is shown in Fig. 3. The effect of dams made all clusters disperse throughout the HRB except for clusters 4 and 5. Cluster 1 was scattered in the southern and eastern regions of the HRB; cluster 2 was mainly distributed along the main stream of the Huai River; most of clusters 3 and 6 were mainly found in the eastern region of the HRB, with a few sites scattered in the western region; clusters 4 and 5, were uniformly concentrated in the central-northern region of the HRB.

The regions of cluster 5 (thick cross in Fig. 3) have been severely polluted and the concentrations of NH_3–N, COD_{Mn}, COD_{Cr}, TP, TN were extremely high. Meanwhile, phytoplankton in these regions made the greatest contributions to the HRB phytoplankton communities due to their highest dominance and biodiversity.

The regions of cluster 4 (thin cross in Fig. 3) were less polluted compared with those of cluster 5. The biodiversity of phytoplankton in these regions was low, which resulted in a lower self-purifying rate in these waters. That in turn makes water quality worse. The emphasis should be laid on improvement of biodiversity and pollutant control in these regions.

In summary, regions of clusters 4 and 5 (within the ellipse in Fig. 3) were severely polluted. Phytoplankton there are highly important to the HRB phytoplankton communities due to the high

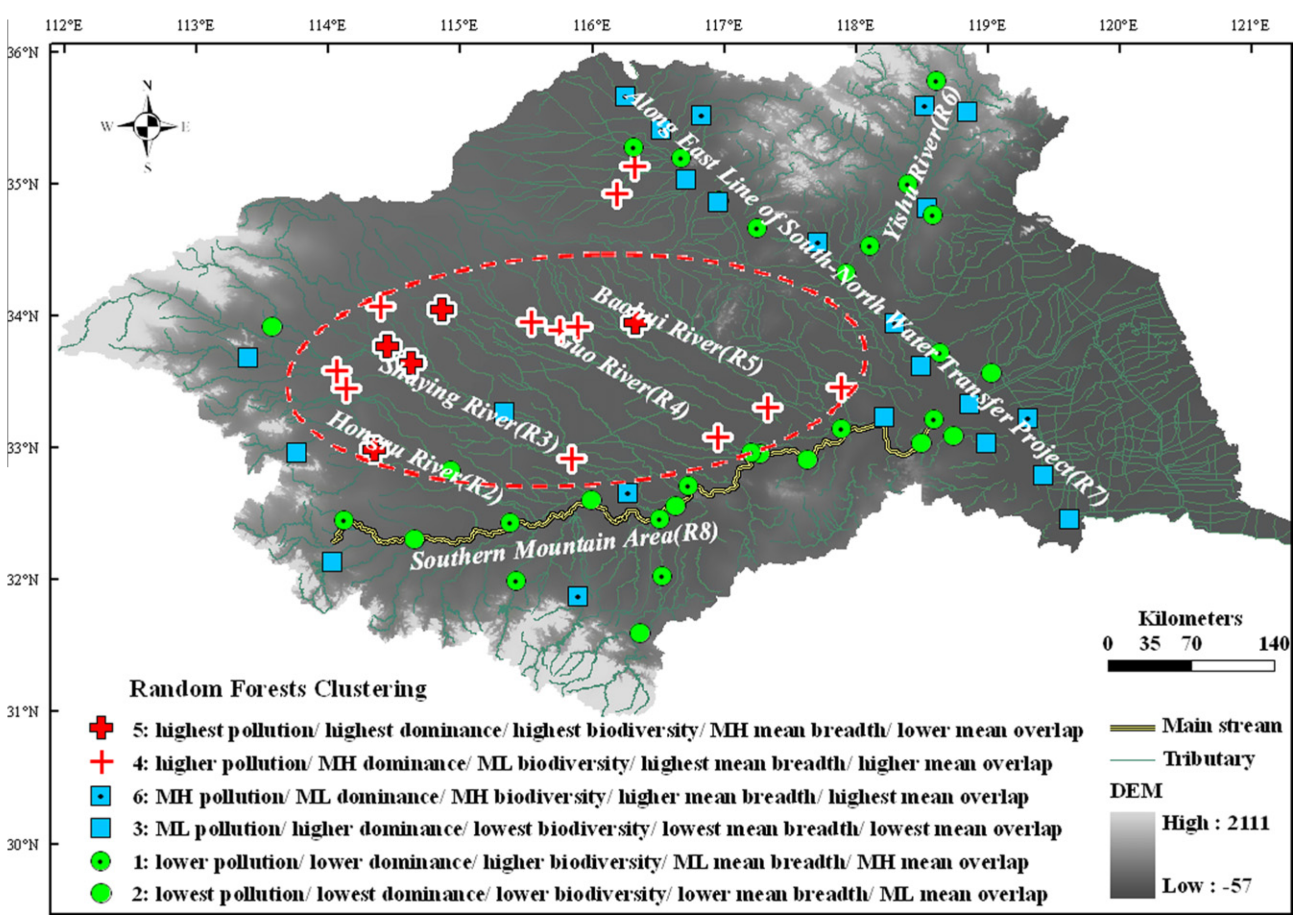

Fig. 3. Spatial random forests clustering in the basin: "MH" and "ML" stand for "middle higher" and "middle lower". DEM means digital elevation model.

dominance and biodiversity in these regions. While ecological restoration measures might cause damage to the habitat and the biota, the relatively high mean breadth values in these regions indicate that the phytoplankton there have strong adaptability to environmental change. Ecological restoration focusing on aquatic environment improvement in the regions of clusters 4 and 5 is, therefore, feasible, and the importance of the phytoplankton there means that these are undoubtedly key regions for future phytoplankton-related ecological restoration.

Several previous studies on the HRB have also found that the northern plain regions of the HRB (approximate the ellipse area in Fig. 3), and especially the Hongru (R2), Shaying (R3) and Guo (R4) Rivers, have the most severe water pollution (Wang and Ongley, 2004; Cheng et al., 2005; Tang et al., 2008; Zhang and Shan, 2008; Xia et al., 2011). These regions have a high population and high levels of industrial activities (e.g. mining of sulfate and chloride minerals), and runoff from agricultural activities (Zhang et al., 2011; Zhao et al., 2008). Zhang et al. (2011) studied the chemistry of these rivers of the HRB, and found that the northern plain regions in the HRB have high concentrations of ions, with the spatial patterns and ionic composition reflecting the intensive human activities in the region. The middle and lower reaches of most rivers have fragile or even unstable aquatic ecosystems, with sub-healthy or unhealthy aquatic habitats (Zhao et al., 2008). Our methods also identified these areas as high-priority sites key sites for restoration, demonstrating that our methods are effective and practical.

We aimed to explore the impact of water pollution on phytoplankton communities after we had identified water pollution as the most important factor influencing the quality of phytoplankton communities. For this purpose, some variables that directly control phytoplankton species composition such as light, silica and residence time were not included within the study. This may result in some uncertainties in results.

The quantity and nature of the required data can limit the effectiveness of methods for identification of key regions in a basin with severe water pollution. Turpie et al. (2002) devised a method for prioritizing South African estuaries on the basis of conservation importance. Estuaries were scored in terms of their size, type and biogeographical zone, habitats and biota (plants, invertebrates, fish and birds). This method considered as many factors as possible and was developed to aid in decision-making regarding the freshwater requirements of estuaries, and in the development of a sound management strategy for estuaries. However, too many data requirements and uncertainties in estimates of the indices limited its application in data-scarce areas. Sánchez-Fernández et al., 2004) identified high-priority areas for conservation using only data on water beetle presence and distribution. However, they did not consider the impact of habitat factors such as water chemistry and other factors controlling phytoplankton species compositions. Filipe et al. (2004) presented a practical way of ranking watercourses for conservation based on the probability of occurrence of species and criteria for rarity, abundance, and endemic value. However, failure to incorporate critical habitat variables with the proper scale may result in incorrect classifications. Similarly, when Posadas and Crisci (2001) set priorities in conservation by using phylogenetic diversity measures, they failed to consider the influences of habitat factors on aquatic species. Therefore these methods are not satisfactory for recognizing key regions in ecologically degenerated area with severe pollution.

In most developing countries and regions, water pollution is the principle factor hindering the restoration of aquatic ecosystems. At the same time, collection of detailed data on factors influencing the growth of aquatic biota is costly. Consequently, such datasets are usually scarce. In those countries/regions, our methods can be easily employed to identify key regions for aquatic ecosystem restoration by using simple critical impact factors.

In this study, some uncertainties in the conclusions may occur because of limited data and because the results are based on only one sampling time. In the process of ecological restoration, regular, long-term monitoring on aquatic ecosystems will be required. Long-term monitoring is likely to increase the precision of prediction of the key regions for rehabilitation.

4. Conclusions

To identify the key regions for future phytoplankton related ecological restoration in the Huai River basin (HRB), China, we sampled phytoplankton and water chemistry in the HRB. Species dominance, biodiversity and ecological niches (niche breadth and niche overlap) at all sampling sites were calculated. Then, the random forests (RFs) clustering approach was used to classify all sampling sites into six clusters.

Analyses of the characteristics of the six clusters showed that two clusters in a severely-polluted region of the Northern Plain area of the HRB were key candidates for restoration. The phytoplankton in these regions were an important component of the whole HRB biota due to their high dominance and biodiversity. Their wide niche breadths indicated that they could tolerate the environmental disruption associated with restoration activities. During phytoplankton-related restoration in the HRB, water temperature, concentration of total phosphorus and chemical oxygen demand can be altered with few adverse effects on phytoplankton communities, while measures increasing Ammonia Nitrogen concentration would be highly detrimental.

Acknowledgements

We acknowledge all reviewers and editors for their valuable advices for this paper. We thank Dr. Anne Colville for her effort in English language correction. This research was supported by Project of the Opening Foundation of the China Institute of Water Resources and Hydropower Research (No. IWHRKF201001), the Natural Science Foundation of China (No. 40971023), the National Key Special Project of Sci-tech for water pollution control and regulation (No. 2009ZX07210-006), the public welfare project of Ministry of Water Resources, China (No. 200801001) and the Post-Doc Science Foundation (No. 20100470022), PR China.

We thank all colleagues from the Water Resources Bureau of the Huai River Committee, from Institute of Geographical Science and Natural Resources Research, Chinese Academy of Sciences, and from the East China Normal University for their support and collaboration in the field investigation in 2008.

References

Abellán, P., Sánchez-Fernández, D., Velasco, J., Millán, A., 2005. Assessing conservation priorities for insects: status of water beetles in southeast Spain. Biol. Conserv. 121 (1), 79–90.

Altaba, C.R., 1990. The last known population of the freshwater mussel *Margaritifera auricularia* (Bivalvia, Unionoida) – a conservation priority. Biol. Conserv. 52 (4), 271–286.

Basille, M., Calenge, C., Marboutin, E., Andersen, R., Gaillard, J.-M., 2008. Assessing habitat selection using multivariate statistics: some refinements of the ecological-niche factor analysis. Ecol. Model. 211 (1–2), 233–240.

Brändle, M., Prinzing, A., Pfeifer, R., Brandl, R., 2002. Dietary niche breadth for central European birds: correlations with species-specific traits. Evol. Ecol. Res. 4, 643–657.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Chen, P.F., Wiley, E.O., McNyset, K.M., 2007. Ecological niche modeling as a predictive tool: silver and bighead carps in North America. Biol. Invasions 9 (1), 43–51.

Chen, X., Bu, Z.J., Wang, S.Z., Li, H.K., Zhao, H.Y., 2009. Niches of seven bryophyte species in Hani Peat land of Changbai Mountains. Chinese J. Appl. Ecol. 20 (3), 574–578 (in Chinese).

Cheng, X.S., Jia, L., Cheng, X.F., 2005. Analysis of natural water chemistry characteristics in the Huai River basin and Shandong Peninsula. Water Resour. Protect. 21, 15–18 (in Chinese).

Crandall, K.A., 1998. Conservation phylogenetics of Ozark crayfishes: assigning priorities for aquatic habitat protection. Biol. Conserv. 84 (2), 107–117.

Daily, G.C., 1999. Developing a scientific basis for managing Earth's life support systems. Conserv. Ecol. 3 (2), 14, http://www.consecol.org/vol3/iss2/art14/.

Daily, G.C., Soderqvist, T., Arrow, K., et al., 2000. The value of nature and the nature of value. Science 289, 395–396.

Domínguez-Domínguez, O., Martínez-Meyer, E., Zambrano, L., De León, G.P., 2006. Using ecological-niche modeling as a conservation tool for freshwater species: live-bearing fishes in central Mexico. Conserv. Biol. 20 (6), 1730–1739.

Dudoit, S., Fridlyand, J., Speed, T.P., 2002. Comparison of discrimination methods for the classification of tumors using gene expression data. J. Am. Stat. Assoc. 97, 77–87.

Falkowski, P.G., 1994. The role of phytoplankton photosynthesis in global biogeochemical cycles. Photosyn. Res. 39, 235–258.

Falkowski, P.G., Barber, R.T., Smetacek, V., 1998. Biogeochemical controls and feedbacks on ocean primary production. Science 281, 200–206.

Ferrier, S., 2002. Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? Syst. Biol. 51, 331–363.

Filipe, A.F., Marques, T.A., Tiago, P., Ribeiro, F., Moreira Da Costa, L., Cowx, I.G., Cowx, I.G., 2004. Selection of priority areas for fish conservation in Guadiana River basin, Iberian Peninsula. Conserv. Biol. 18 (1), 189–200.

Foulon, E., Not, F., Jalabert, F., Cariou, T., Massana, R., Simon, N., 2008. Ecological niche partitioning in the picoplanktonic green alga *Micromonas pusilla*: evidence from environmental surveys using phylogenetic probes. Environ. Microbiol. 10 (9), 2433–2443.

Friberg, M., Bergman, M., Kullberg, J., Wahlberg, N., Wiklund, C., 2008. Niche separation in space and time between two sympatric sister species – a case of ecological pleiotropy. Evol. Ecol. 22 (1), 1–18.

Gao, C., Gemmer, M., Zeng, X., Liu, B., Su, B., Wen, Y., 2010. Projected streamflow in the Huaihe River Basin (2010–2100) using artificial neural networks (ANN). Stoch. Environ. Res. Risk. Assess. 24 (5), 685–697.

Giller, P.S., Hillebrand, H., Berninger, U.G., et al., 2004. Biodiversity effects on ecosystem functioning: emerging issues and their experimental test in aquatic environments. Oikos 104 (3), 423–436.

Hardin, J., Rocke, D.M., 2004. Outlier detection in the multiple cluster setting using the minimum covariance determinant estimator. Comput. Stat. Data Anal. 44 (4), 625–638.

Horvath, S., Shi, T., Shai, R.M., Chen, C., Nelson, S., 2007. Statistical Methods Supplement and R software tutorial: Gene Filtering with a Random Forest Predictor. <http://www.genetics.ucla.edu/labs/horvath/RandomForestScreening/>.

Hurlbert, S.H., 1978. The measurement of niche overlap and some relatives. Ecology 59 (1), 67–77.

Hutchinson, G.E., 1957. Concluding remarks. Cold Spring Harb. Symp. 22, 415–427.

Irfan-Ullah, M., Amarnath, G., Murthy, M.S.R., Peterson, A.T., 2007. Mapping the geographic distribution of *Aglaia bourdillonii* Gamble (Meliaceae), an endemic and threatened plant, using ecological niche modeling. Biodivers. Conserv. 16 (6), 1917–1925.

Jehle, R., Bouma, P., Sztatecsny, M., et al., 2000. High aquatic niche overlap in the newts *Triturus cristatus* and *T. marmoratus* (Amphibia, Urodela). Hydrobiologia 437, 149–155.

Jiang, W.X., Fu, X.C., Tang, T., Cai, Q.H., 2009. Community structure and niche of macroinvertebrates in the Xiangxi River in Hubei, China. Chinese J. Appl. Environ. Biol. 15 (3), 337–341 (in Chinese).

Kagalou, I.I., Kosiori, A., Leonardos, I.D., 2010. Assessing the zooplankton community and environmental factors in a Mediterranean wetland. Environ. Monit. Assess. 170 (1–4), 445–455.

Levins, R., 1968. Evolution in Changing Environments: Some Theoretical Explorations. Princeton University Press, Princeton.

Liu, C.M., Zhao, C.S., Xia, J., Sun, C.L., Wang, R., Liu, T., 2011. An instream ecological flow method for data-scarce regulated rivers. J. Hydrol. 398, 17–25.

McNyset, K.M., 2005. Use of ecological niche modelling to predict distributions of freshwater fish species in Kansas. Ecol. Freshwater Fish 14 (3), 243–255.

Mehrian, S.R., Chen, C.D., Shi, T., Horvath, S., Nelson, S.F., Reichardt, J.K.V., Sawyers, C.L., 2007. Insulin growth factor-binding protein 2 is a Biomarker for PTEN Status and PI3K/Akt Pathway Activation in Glioblastoma and Prostate Cancer. Proc. Natl. Acad. Sci. 104 (13), 5563–5568.

Pearce, J., Lindenmayer, D., 1998. Bioclimatic analysis to enhance reintroduction biology of the endangered helmeted honeyeater (*Lichenostomus melanops cassidix*) in southeastern Australia. Restor. Ecol. 6, 238–243.

Pérez-Losada, M., Jara, C.G., Bond-Buckup, G., Crandall, K.A., 2002. Conservation phylogenetics of Chilean freshwater crabs *Aegla* (Anomura, Aeglidae): assigning priorities for aquatic habitat protection. Biol. Conserv. 105 (3), 345–353.

Peterson, A.T., Nakazawa, Y., 2008. Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*. Global Ecol. Biogeogr. 17 (1), 135–144.

Peterson, A.T., Papes, M., Eaton, M., 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography 30 (4), 550–560.

Pianka, E.R., 1974. Niche overlap and diffuse competition. Proc. Natl. Acad. Sci. 71, 2141–2145.

Posadas, P.M.D., Crisci, J.V., 2001. Using phylogenetic diversity measures to set priorities in conservation: an example from southern South America. Conserv. Biol. 15, 1325–1334.

Quero, J.L., Gomez-Aparicio, L., Zamora, R., Maestre, F.T., 2008. Shifts in the regeneration niche of an endangered tree (*Acer opalus* ssp. *granatense*) during ontogeny: using an ecological concept for application. Basic Appl. Ecol. 9 (6), 635–644.

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org/>.

Raxworthy, C.J., Ingram, C.M., Rabibisoa, N., Pearson, R.G., 2007. Applications of ecological niche modeling for species delimitation: a review and empirical evaluation using day geckos (*Phelsuma*) from Madagascar. Syst. Biol. 56 (6), 907–923.

Sánchez-Fernández, D., Abellán, J., Velasco, J., Millán, A., 2004. Selecting areas to protect the biodiversity of aquatic ecosystems in a semiarid Mediterranean region using water beetles. Aquat. Conserv. 14 (5), 465–479.

Shafi, I., Ahmad, J., Shah, S.I., Ikram, A.A., Khan, A.A., Bashir, S., 2010. Validity-guided fuzzy clustering evaluation for neural network-based time-frequency reassignment. EURASIP J. Adv. Sig. Pr., 1–14.

Shannon, C.E., Weaver, W., 1949. The Mathematical Theory of Communication. The University of Illinois Press, Urbana, 117 pp.

Shi, T., Seligson, D., Belldegrun, A.S., Palotie, A., Horvath, S., 2005. Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma. Mod. Pathol. 18, 547–557.

Sigee, D.C., 2005. Freshwater Microbiology: Biodiversity and Dynamic Interactions of Microorganisms in the Aquatic Environment. John Wiley & Sons Ltd, England.

Smith, E.P., 1982. Niche Breadth, Resource Availability and Statistical Inference. SIMS Technical Report No. 8, Biomathematics Group, University of Washington.

Solano, E., Feria, T.P., 2007. Ecological niche modeling and geographic distribution of the genus *Polianthes* L. (Agavaceae) in Mexico: using niche modeling to improve assessments of risk status. Biodivers. Conserv. 16 (6), 1885–1900.

Spellerberg, I.F., Fedor, P.J., 2003. A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener' Index. Global Ecol. Biogeogr. 12, 177–179.

Tang, Q.Y., Feng, M.G., 2007. DPS Data Processing System: Experimental Design, Statistical Analysis, and Data Mining. Science Press, Beijing (in Chinese).

Tang, L.H., Zhang, S.C., Lu, X.B., Liu, M.H., 2008. Estimation of agricultural non-point source pollution loads of the Huai River basin in the East Line of South North Water Transfer Project. J. Agric. Environ. Sci. 27, 1437–1441 (in Chinese).

Thompson, K., Gaston, K.J., 1999. Range size, dispersal and niche breadth in the herbaceous flora of central England. Ecology 87, 150–155.

Thorn, J.S., Nijman, V., Smith, D., Nekaris, K.A.I., 2009. Ecological niche modelling as a technique for assessing threats and setting conservation priorities for Asian slow lorises (Primates: *Nycticebus*). Divers. Distrib. 15 (2), 289–298.

Turpie, J.K., Adams, J.B., Joubert, A., et al., 2002. Assessment of the conservation priority status of South African estuaries for use in management and water allocation. Water SA 28 (2), 191–206.

Vitousek, P.M., Mooney, H.A., Lubchenco, J., et al., 1997. Human domination of Earth's ecosystems. Science 277, 494–499.

Waltari, E., Guralnick, R.P., 2009. Ecological niche modelling of montane mammals in the Great Basin, North America: examining past and present connectivity of species across basins and ranges. J. Biogeogr. 36 (1), 148–161.

Wang, C.F., Ongley, E.D., 2004. Transjurisdictional water pollution management: the Huai River example. Water Int. 29, 290–298.

Wang, G.S., Xia, J., 2010. Improvement of SWAT2000 modelling to assess the impact of dams and sluices on streamflow in the Huai River basin of China. Hydrol. Process. 24 (11), 1455–1471.

Wiley, E.O., McNyset, K.M., Peterson, A.T., et al., 2003. Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. Oceanography 16, 120–127.

Xia, J., Zhang, Y.Y., Zhan, C.S., Ye, A.Z., 2011. Water quality management in China: the case of the Huai River basin. Int. J. Water Resour. D 27 (1), 167–180.

Zhang, H., Shan, B.Q., 2008. Historical records of heavy metal accumulation in sediments and the relationship with agricultural intensification in the Yangtze – the Huai River region, China. Sci. Total Environ. 399, 113–120.

Zhang, Y.Y., Xia, J., Liang, T., Shao, Q.X., 2010. Impact of water projects on river flow regimes and water quality in Huai River Basin. Water Resour. Manage. 24 (5), 889–908.

Zhang, L., Song, X.F., Xia, J., et al., 2011. Major element chemistry of the Huai River basin, China. Appl. Geochem. 26 (3), 293–300.

Zhao, C.S., Xia, J., Wang, G.S., et al., 2008. Evaluation and analysis on aquatic ecology and environmental quality of Huai River basin. Chinese J. Environ. Eng. 12 (2), 1698–1704 (in Chinese).

Zhao, C.S., Sun, C.L., Xia, J., et al., 2010. An impact assessment method of dam/sluice on instream ecosystem and its application to the Bengbu Sluice of China. Water Resour. Manage. 24, 4551–4565.

Zhao, C.S., Liu, C.M., Xia, J., et al., 2011. Zoobenthos' ecological niche clustering for key regions recognization in ecological restoration of the Huai River, China. Ecohydrology. Revised.