Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :
Chuanbo GUO
Le mardi 27 mai 2014

Titre :

Modélisation des effets des changements climatiques et des activités anthropiques sur les assemblages des poissons des lacs en Chine

---

ED SEVAB : Écologie, biodiversité et évolution

Unité de recherche :

Laboratoire Évolution et Diversité Biologique (UPS, France), Institute of Hydrobiology (CAS, China)

Directeur(s) de Thèse :

Prof. Zhongjie LI (Institute of Hydrobiology, Wuhan, China)
Prof. Sovan LEK(Université de Paul Sabatier, Toulouse, France)

Rapporteurs :

Prof. Erik Jeppesen (Aarhus University, Denmark)
Prof. Zuogang Peng (Southwest University,China)

Autre(s) membre(s) du jury :

Prof. Yan Wang (Zhejiang University, China)
Prof. Qiwei Wei (Yangtze River Fisheries Research Institute, China)
Prof. Jiashou Liu (Institute of Hydrobiology, China)
Prof. Zhan Yin (Institute of Hydrobiology, China)

THESE

En vue de l’obtention du

DOCTORAT DE L’UNIVERSITÉ DE TOULOUSE

Délivré par :

UNIVERSITE TOULOUSE 3 PAUL SABATIER (UT3 PAUL SABATIER)

Présentée et soutenue par :

CHUANBO GUO

Modélisation des effets des changements climatiques et des activités anthropiques sur les assemblages des poissons des lacs en Chine

Directeur(s) de Thèse :

Prof. Zhongjie LI (Institute of Hydrobiology, Wuhan, China)
Prof. Sovan LEK(Université de Paul Sabatier, Toulouse, France)

Jury

Prof. Erik Jeppesen (Aarhus University, Denmark)	Rapporteur
Prof. Zuogang Peng (Southwest University,China)	Rapporteur
Prof. Sovan Lek (Toulouse University, France)	Co-directeur
Prof. Zhan Yin (Institute of Hydrobiology, China)	Examinateur
Prof. Jiashou Liu (Institute of Hydrobiology, China)	Examinateur
Prof. Yan Wang (Zhejiang University, China)	Examinateur
Prof. Qiwei Wei (Yangtze River fisheries research, China)	Examinateur

Table of Contents

Part Ⅱ: Publications

[1] Chuanbo Guo, Shaowen Ye, Sovan Lek, Jiashou Liu, Tanglin Zhang, Jin Yuan & Zhongjie Li. The need for improved fishery management in a shallow macrophytic lake in the Yangtze River basin: evidence from the food web structure and ecosystem analysis. *Ecological Modelling, 267 (2013) 138–147*

[2] Chuanbo Guo, Sovan Lek, Shaowen Ye, Wei Li, Jiashou Liu, Tanglin Zhang, Jing Yuan & Zhongjie Li. Large-scale patterns of fish species distribution and assemblage: a novel multi-species modeling approach. *Freshwater biology* (Under review)

[3] Chuanbo Guo, Sovan Lek, Wei Li, Jiashou Liu, Zhang Tanglin, Shaowen Ye & Zhongjie Li. Predicting fish species richness and assemblages in Chinese lakes using Multivariate Regression Trees. *Biodiversity and Conservation* (Under review)

[4] Chuanbo Guo, Sovan Lek, Shaowen Ye, Wei Li, Jiashou Liu & Zhongjie Li. Uncertainty in ensemble modelling of large-scale species distribution: effects from the species characteristics and model techniques. *Ecological Modelling* (Under review)

ACKNOWLEDGMENTS

It would not have been possible to finish this dissertation without the help and support of the kind people surround me, to only some of whom it is possible to give particular mention here.

My first and sincere appreciation goes to Prof. Zhongjie Li and Prof. Sovan Lek, my two respected supervisors for all I have learned and gained from them and for their continuous help and support in all stages of this thesis. I would like to thank Prof. Zhongjie Li for being an open person to ideas, and for encouraging and helping me to shape my interest and ideas. His attitude to research inspired me to continue to a PhD program and want to be a member of the academic family. I would also like to express my deep gratitude and respect to Prof. Sovan lek whose advices and insight was invaluable to me. Just as somebody else mentioned that, Sovan taught me not only how to make an ecological model, but also how to model my whole life. Great appreciations should be given to them again for providing me the chance to study abroad.

My greatest and deepest gratitude goes also to Prof. Jiashou Liu, for providing invaluable suggestions on my research work and experiment. Special thanks also go to Dr. Ye Shaowen who helps me a lot in writing and revising my manuscripts. Thanks to Prof. Tanglin Zhang, Dr. Li Wei for sharing experience in field work and providing constructive suggestions on my research.

I would like to express my great appreciations to Associate Prof. Sithan Lek who take care my life during my stay in France, thanks a million for your delicious food which always with different style all over the world each time, up until now, never repeat.

Special thanks go to Prof. Gaël GRENOUILLET for discussing with me and explaining douts on ecological modeling; to Dr. Loïc TUDESQUE for teaching me Arcgis; to Dr. Matieu CHEVALIER for solving the bug in data analysis; to Prof.

Christina LAUZERAL for providing the convenience in using the computer. I would like to express my great thanks to Prof. Sébastien BROSSE, Dr. Simon BLANCHET and Dr. Julien CUCHEROUSSET for their kindly assistance and suggestions on my research proposal. Without all of your help, I couldn’t continue my research work in a new field in ecological modeling.

I’m greatly indebted to Prof. Xie Shouqi, Prof. Yu Yuhe, Prof. Xie Songguang, Prof. Wang Hongzhu, Prof. Chen Yifeng, associate Prof. Zhu Xiaoming for their constructive suggestions and instructions on my research. Thanks to Miss Yang Yunxia, associate Prof. Wang Haijun for their help in my experiment. Thanks to Mr. Chen Xinnian and Miss Yuan Jin for creating quite a suitable environment for our research.

My greatest appreciation and friendship goes to my closest friends and colleagues both in “Laboratory of fisheries ecology” at IHB, China (they are Zeng Wentao, Cheng Lin, Guo Zhiqiang, Lin Mingli, Zhu Fengyue, Tang Jianfen, Zhu Tingbing, Zhang Lihong, Du Xue, Zhang Chaowen, Yang Zhanwei, Wang Qidong, Xia Yuguo, Ren Peng, Cui Fangtian, Qin Xiangchao, Tang Lijun, Fan Hourui, Chen Sibao, Li Chang, Ren Yan, Cai Xingwei, Wang Jingya, Cheng Houjun and Lian Yuxi) and “Laboratory of Evolution, Diversity and Biology” at UPS, France (Dr. Gaël GRENOUILLET, Prof. Sébastien BROSSE, Dr. Simon BLANCHET, Dr. Loïc TUDESQUE, Dr. Lorenza CONTI, Dr. Christine LAUZERAL, Lise COMTE, Nicolas CHARPIN, etc.). Special thanks should owe to Dr. Zhao Tian and his wife Dr. Guo Jianwei who were always a great support in all my struggles and frustrations in my new life and studies in France. I am so lucky that I can work with these two big families!

At the end, I would like to thank the Graduate Student Affairs Division of IHB (Mr. Xi FENG, Mrs. Kefeng LIU, Mrs. Mei ZHA and Mrs.Caipai LIAO) and the secretariat of EDB (Mrs. Linda JALABERT, Mrs. Nicole HOMMET and Mr. Frédéric MAGNÉ). I wish to thank Mrs. Dominique PANTALACCI for her generous helps with the registration, TOEIC test and so on.

Last but not least, I would like to express my great appreciations to my parents, without their support I couldn’t move on so fast. I sincerely thank my beloved girlfriend who accompany with me to overcome difficulties I met, she encourages me to progress. I’m quite sure that our lives will be better and better.

I was concerned that I might forget to mention lots of you in this short acknowledgement, but please accept my faithfully wishes, to all of you!

This research was financially supported by Grant Nos. 200903048 and 20130305 from the “Special Fund for Agro-scientific Research in the Public Interest” of China, Nos. 30830025 and 30900182 from the National Natural Science Foundation of China. We are grateful to the China Scholarship Council (CSC) for their financial support of my study abroad.

Faithfully yours

Chuanbo Guo

Toulouse

01/01/2014

Résumé

La connaissance de la distribution spatiale des espèces et des communautés dans les écosystèmes est un préalable indispensable à la compréhension du fonctionnement et les processus des écosystèmes ainsi que les questions de conservation et d'aménagement du territoire. Dans le contexte des changements globaux, le changement climatique et les activités anthropiques sont reconnus comme des déterminants majeurs de la diversité et des patrons de distribution des poissons, affectant ainsi les propriétés et la structure des écosystèmes aquatiques. Cependant, très peu d’efforts ont été entrepris pour étudier la diversité et le fonctionnement de ces écosystèmes lacustres en Chine. Dans cette thèse, nous contribuons à mettre en évidence les effets des changements climatiques et des activités anthropiques sur les patrons de diversité et de distribution des poissons, ainsi que sur le fonctionnement des écosystèmes en nous appuyant sur des approches de modélisation écologique.

Dans le premier temps, nous avons déterminé les distributions et les assemblages globales de 425 espèces et sous-espèces de poissons dans 135 lacs distribués à travers le territoire chinois en utilisant une nouvelle approche multi-espèce ajustée par l’arbre multi-variable de régression et connu sous le nom de MRT. Cinq assemblages de poissons ont pu être définis par la classification contrainte avec un total de 107 espèces indicatrices identifiées. Nous avons montré une différence significative de la diversité spécifique pour chaque assemblage : la diversité spécifique des poissons dans le plateau étant significativement plus faible que celle des lacs des plaines. Cependant, la diversité de l’assemblage de l’ensemble des plateaux est plus importante que celle d'autres régions. Par ailleurs, nos résultats indiquent que l’altitude, la température minimale du mois le plus froid, la gamme de température annuelle et les précipitations du mois le plus sec sont des facteurs important dans la détermination des assemblages et des patrons de distribution des poissons des lacs chinois.

Dans un 2$^{\text{ème}}$ temps, le modèle MRT a été utilisé pour prédire la richesse

spécifique et la distribution des espèces de poissons sous l'influence du changement climatique dans le but d'améliorer la gestion et la conservation des poissons en Chine. Nos résultats montrent que MRT est un modèle fiable et idéal pour prédire la communauté multi-spécifique de poissons. Au niveau de la composition spécifique, l'altitude apparait comme le meilleur des prédicteurs, suivi par les précipitations au cours du mois le plus sec, la gamme de température annuelle et la température annuelle moyenne. Pour la richesse spécifique, les précipitations au cours du mois le plus sec, la température maximale du mois le plus chaud et l'aire des lacs sont des facteurs majeurs pour prédire les patrons de distribution et de richesse des poissons.

Dans un 3ème temps, nous examinons la capacité et l'incertitude associé aux modèles d'ensemble (c'est à dire à la combinaison des prédictions issues de diffrents modèles de distribution d'espèces (SDM)) pour prédire la distribution et la diversité des espèces de poissons. L'impact potentiel de 2 sources d'incertitude a été démontré : les caractéristiques des espèces (c'est à dire la prévalence des espèces, l'altitude, les gammes de températures et de précipitations) et les techniques de modélisation (méthodes de calibration et d'évaluation). Nos résultats montrent que les prédictions issues d'un simple modèle de distribution (SDM) est très variable et même douteuse pour l'ensemble des espèces considérée alors qu'une approche d'ensemble donne de meilleures prédiction. Nous avons aussi montré qu'il n'y a pas d'influence significative des méthodes d'évaluation sur les sorties du modèle. Nous montrons que les caractéristiques des espèces, la prévalence des espèces, l'altitude et la gamme de précipitations pourraient fortement affecter les résultats des SDMs, alors que la gamme de températures n'a pas d'influence significative. Finalement, nous avons vérifié l'hypothèse selon laquelle la distribution des espèces à faible aire de distribution peut être prédite avec plus de précision que la distribution des espèces à large échelle de distribution.

Finalement, l'étude a été focalisée sur l'évaluation des effets des activités anthropiques sur la structure et les fonctions des écosystèmes lacustres peu profond riche en macrophytes (lac Bao'an, localisé dans la partie médiane du Yangtze) en utilisant le modèle ECOPATH. Les résultats montrent que les espèces de poissons

commercialisables souffrent d'une très forte pression de pêche, tandis que les ressources fourragères comme les algues, les plantes immergés et les mollusques ne présentes pas de signe de surexploitation. En outre, nos résultats montrent que l'écosystème du lac Bao'an est un système "mûr" selon la théorie d’Odum. Cependant, comparé aux autres écosystèmes lacustres, le lac Bao’an, comme beaucoup d’autres lacs en Chine, montrent une très faible valeur de CI (Connectance Index), de FCI (Finn’sCycling Index) et de SOI (System Omnivory Index), indiquant que les fonctions de ces écosystèmes lacustres chinois ont tendance à être plus simple et plus linéaire que les écosystèmes lacustres d'autres pays. En conséquence, notre étude témoigne d'un besoin urgent d’orientation et de management des activités anthropiques, principalement le repeuplement et la pêcherie traditionnelle.

Mots-clés: Modèle de Distribution d’Espèces, Ecopath with Ecosim, Assemblages de Poissons, Patron de Distribution, Prédiction, Propriété et Structure de l’Ecosystème, Lacs, Chine

Abstract

Knowledge of the spatial distribution of species and communities in ecosystems is an essential prerequisite for the understanding of ecosystem functioning and processes as well as conservation and spatial planning issues. During the last several decades, in the context of global change, climate change and anthropogenic activities have long been acknowledged as the two main determinants which drive the fish diversity and distributions patterns, and ultimately affect the aquatic ecosystem properties and structure. However, up until now, very few efforts aimed at the fish diversity and ecosystem in the lakes across China. Consequently in the present study, we contribute to highlight the effects of climate change and anthropogenic activities on fish diversity and distribution patterns as well as the ecosystem properties with the approach of several ecological modelling.

Specifically, we first build the global perspective on the fish distribution and assemblage patterns for a total of 425 fish species (subspecies) in 135 lakes across China using a novel multi-species approach fitted by the Multivariate Regression Tree (MRT). Five fish assemblages were defined by the constrained clustering, 107 indicator species were thus identified. Species diversity showed significantly differences among each assemblage: fish species richness in plateau lakes was significantly lower than plain lakes; however the diversity of the whole assemblage in plateaus was higher than other regions. Altitude, minimum temperature of the coldest month, annual temperature range and precipitation during the driest month were found to be the most important determinants affecting fish assemblages and distribution patterns in Chinese lakes.

Then, MRT model was used to predict both species richness and species distribution in order to improve the management and conservation of fish species in China. Our results showed that MRT is a reliable and ideal community-based predictive technique for multi-species prediction. At the species composition level, altitude was the main determinant for the prediction, followed by precipitation of the

driest month, temperature annual range and annual mean temperature. While at the richness level, precipitation of driest month, maximum temperature of warmest month and lake area were the main drivers for the prediction of the fish species richness pattern.

Thirdly, we examine the capacity and uncertainty of ensemble modelling in predicting fish species distribution and diversity. Potential impacts from two main kinds of uncertainty sources were thus considered: species characteristics (contained species prevalence, altitude range, temperature range and precipitation range) and model techniques (calibration technique and evaluation technique). Finally, our results highlight that predictions from single SDM were so variety and unreliable for all species while ensemble approaches could yield more accurate predictions; we also found that there was no significant influence on the model outcomes from the evaluation measures; we emphasized that species characteristics as species prevalence, altitude range size and precipitation range size would strongly affect the outcomes of SDMs, but temperature range size didn’t show a significantly influence; our findings finally verified the hypothesis that species distributed with a smaller range size could be more accurately predicted than species with large range size to be plausible in aquatic ecosystems.

Lastly, a case study focused on evaluating the lake ecosystem properties and foodweb structure as well as the effects in a typical shallow macrophytic lake (Bao’an Lake, distributed in the middle reaches of the Yangtze River basin), using the Ecopath model. Finally, the results showed that all the commercial fish suffered from high fishing pressure while forage resources such as attached algae, submerged plants and molluscs were not fully utilized. Moreover, we highlight that the Bao’an Lake ecosystem was a mature system according to Odum’s theory. However when compared with some other lake ecosystems, the Bao’an Lake ecosystem, as well as some China lake ecosystems, showed extremely low values of CI (Connectance Index), FCI (Finn’s Cycling Index) and SOI (System Omnivory Index), indicating that the ecosystem functions and food web structure of these Chinese lake tended to be simpler and linear than lake ecosystems in other countries. Consequently, this study

indicated an urgent need for the adjustment and management of artificial fishery stocking in such type of lakes.

Our present study have pictured the global perspective of lake fish diversity and distribution patterns in China, defined the main determinants, and examined the potential effects of climate change and anthropogenic activities on fish diversity and ecosystem properties. Our results will benefit the conservation and management of fish resources, biodiversity, as well as the lake ecosystems all over the world.

Key words: Species Distribution Model, Ecopath with Ecosim, Fish assemblages, Distribution patterns, Predicting, Ecosystem properties and structure, Lakes, China

Part I: Synthesis

1. Introduction

1.1Global climate change and biodiversity

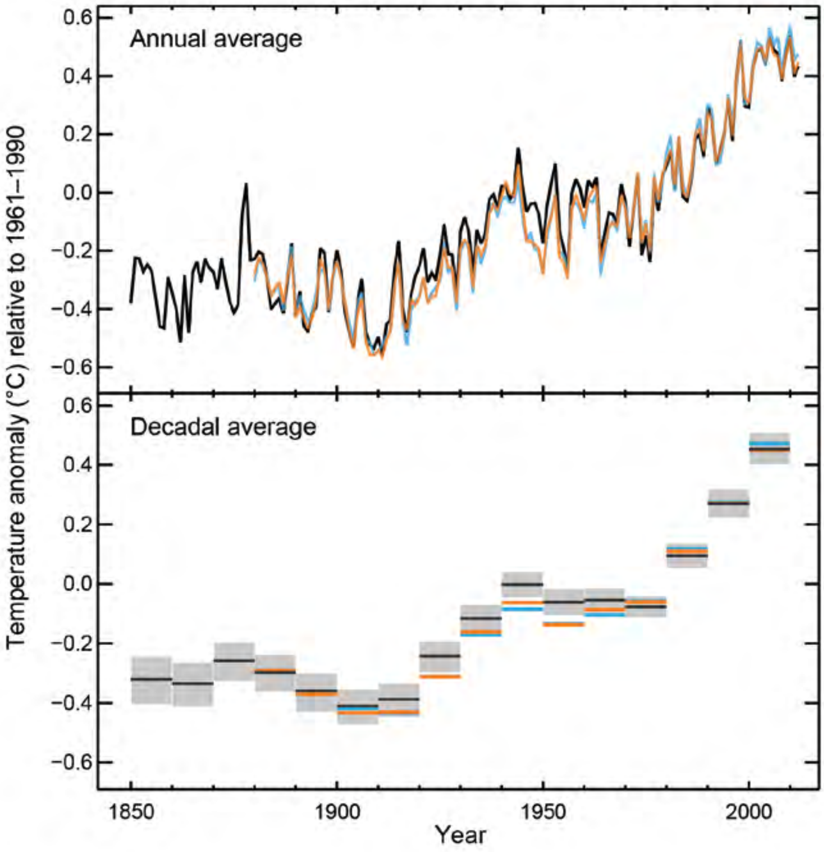
1.1.1 Climate change

Increasingly publications consider new evidence of climate change based on many independent scientific analyses from observations of the climate system, paleoclimate archives, theoretical studies of climate processes and simulations using climate models. Reports from the IPCC 2013 showed that warming of the climate system is unequivocal, and since the 1950s, many of the observed changes are unprecedented over decades to millennia. The atmosphere and ocean have warmed, the amounts of snow and ice have diminished, sea level has risen, and the concentrations of greenhouse gases have increased. According to the report, natural and anthropogenic substances and processes that alter the Earth's energy budget are drivers of climate change (IPCC, 2013).

Specifically, over the period 1880 to 2012, there has been a consistent, large-scale warming of both the land and ocean surface. The globally averaged combined land and ocean surface temperature show a warming of 0.85 (0.65 to 1.06) °C, the total increase between the average of the 1850–1900 period and the 2003–2012 period is 0.78 (0.72 to 0.85) °C, based on the single longest dataset available (Fig.1a), While for the longest period when calculation of regional trends is sufficiently complete (1901 to 2012), almost the entire globe has experienced surface warming (Fig.1b).

Increasing global mean surface temperature is very likely to lead to changes in precipitation and atmospheric moisture because of changes in atmospheric circulation, a more active hydrological cycle, and increases in the water-holding capacity throughout the atmosphere. Confidence in precipitation change averaged over global land areas since 1901 is low prior to 1951 and medium afterwards. Averaged over the mid-latitude land areas of the Northern Hemisphere, precipitation has increased since

(a) Observed globally averaged combined land and ocean surface temperature anomaly 1850–2012



(b) Observed change in surface temperature 1901–2012

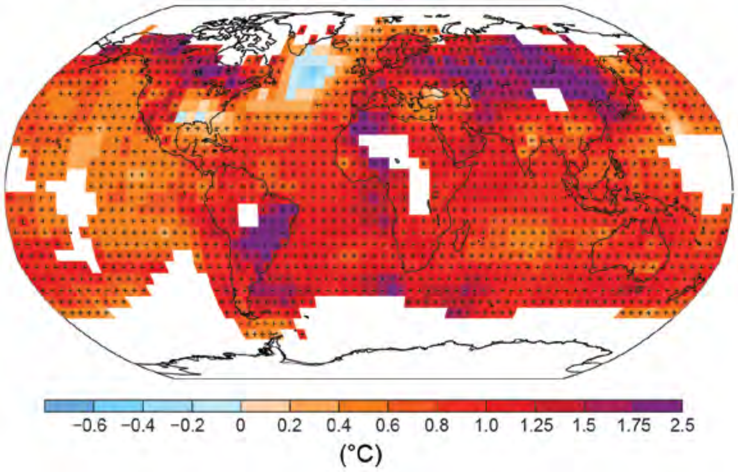


Fig.1 (**a**) Observed global mean combined land and ocean surface temperature anomalies, from 1850 to 2012 from three data sets. Top panel: annual mean values. Bottom panel: decadal mean values including the estimate of uncertainty for one dataset (black). Anomalies are relative to the mean of 1961−1990. (**b**) Map of the observed surface temperature change from 1901 to 2012 derived from temperature trends determined by linear regression from one dataset (orange line in panel a). *Source: (IPCC, 2013)*

1901 (medium confidence before and high confidence after 1951) (Fig.2). Precipitation has very likely increased during the 20th century by 5 to 10% over most mid- and high latitudes of Northern Hemisphere continents, but in contrast rainfall has likely decreased by 3% on average over much of the subtropical land areas (IPCC, 2013).

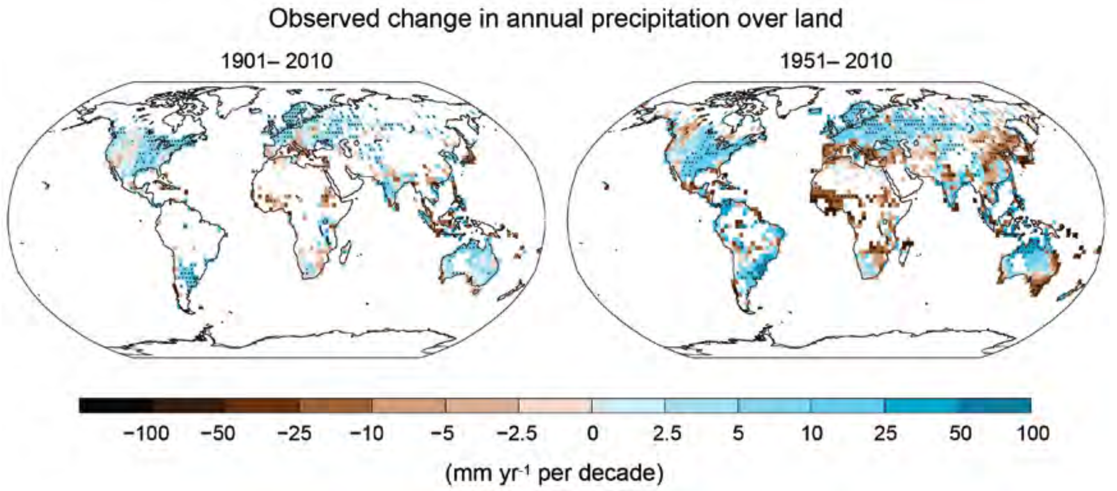


Fig.2 Maps of observed precipitation change from 1901 to 2010 and from 1951 to 2010. Source: *(IPCC, 2013)*

Moreover, there is very high confidence that the rate of sea level rise since the mid-19th century has been larger than the mean rate during the previous two millennia. Over the period 1901 to 2010, global mean sea level rose by 0.19 (0.17 to 0.21) m. The atmospheric concentrations of carbon dioxide, methane, and nitrous oxide have increased to levels unprecedented in at least the last 800,000 years. Carbon dioxide concentrations have increased by 40% since pre-industrial times, primarily from fossil fuel emissions and secondarily from net land use change emissions. The ocean has absorbed about 30% of the emitted anthropogenic carbon dioxide, causing ocean acidification (IPCC, 2013).

While in the future global and regional climate change, projections of changes in the climate system are made using a hierarchy of climate models ranging from simple climate models, to models of intermediate complexity, to comprehensive climate models, and Earth System Models. These models simulate changes based on a set of scenarios of anthropogenic forcing. A new set of scenarios, the Representative Concentration Pathways (RCPs), was used for the new climate model simulations carried out under the framework of the Coupled Model Intercomparison Project Phase 5 (CMIP5) of the World Climate Research Programme.

Projections of temperature and precipitation for the end of the 21st century (2081 – 2100) are given relative to 1986 – 2005, increase of global mean surface temperatures for 2081–2100 relative to 1986–2005 is projected to likely be in the ranges derived from the concentration-driven CMIP5 model simulations, that is, 0.3°C to 1.7°C (RCP2.6), 1.1°C to 2.6°C (RCP4.5), 1.4°C to 3.1°C (RCP6.0), 2.6°C to 4.8°C (RCP8.5).Warming will continue beyond 2100 under all RCP scenarios except RCP 2.6 (Fig.3a). While for the projections of global precipitation, extreme precipitation events over most of the mid-latitude land masses and over wet tropical regions will very likely become more intense and more frequent by the end of this century, as global mean surface temperature increases (Fig.3b).

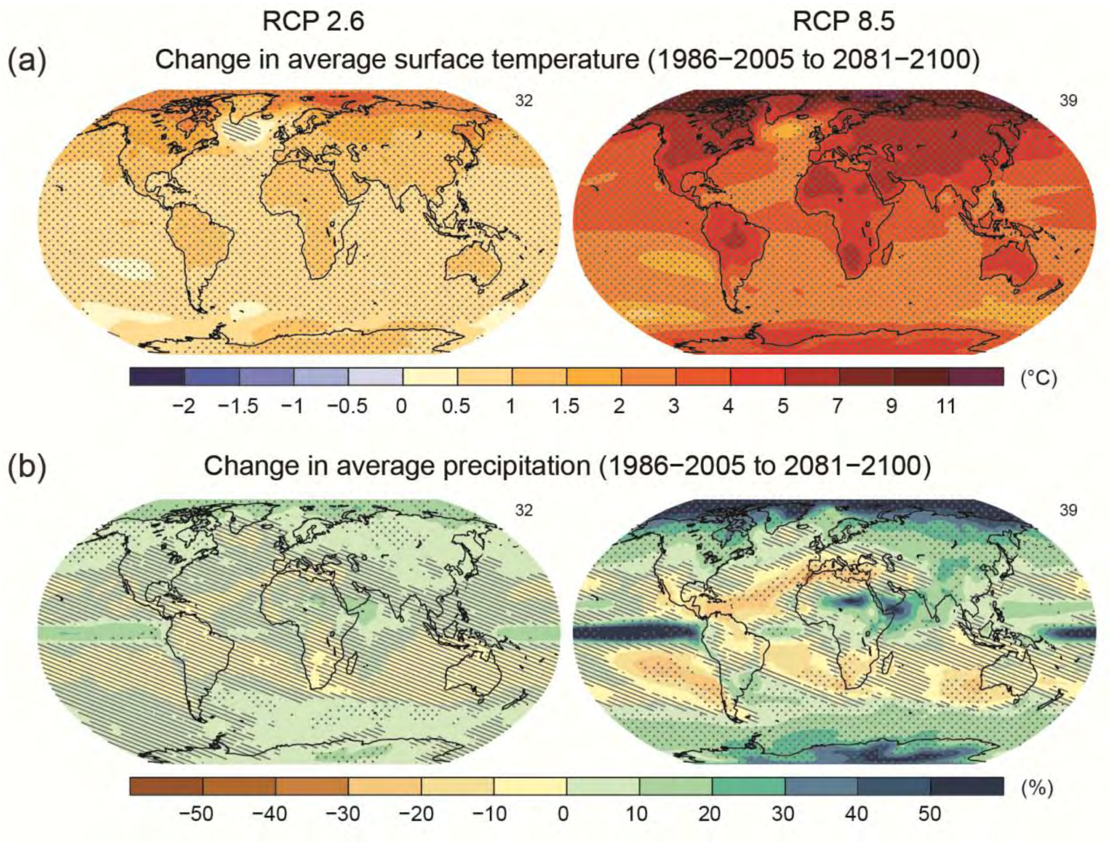


Fig.3 Maps of CMIP5 multi-model mean results for the scenarios RCP2.6 and RCP8.5 in 2081–2100 of (a) annual mean surface temperature change, (b) average percent change in annual mean precipitation.

Source: (IPCC, 2013)

1.1.2 Impacts of climate change on freshwater biodiversity and ecosystem

There is ample evidence that global climate change will irreversibly affect the biodiversity and ecosystem, especially the freshwater ecosystems which were always considered as one of the most vulnerable ecosystem (Millenium Ecosystem Assessment, 2005). However, knowledge of the biodiversity of freshwater ecosystems is still very incomplete, but declines in biodiversity are thought to be far greater in freshwater than in the most affected terrestrial ecosystems (Dudgeon et al., 2006).

Fish assemblage and diversity, were always considered as the most important indicator for freshwater ecosystem, have been frequently used as the research hotspot in the context of climate change recently. Abell et al. (2008) declared that fish are directly affected by changes in temperature since that the richness of freshwater fish communities showed a significantly decrease from the tropical to the polar region. Moreover, Jeppesen et al. (2010) compared the fish assemblages in 94 shallow lakes in Denmark (cold temperate), Belgium and the Netherlands (temperate) and Southern Spain (Mediterranean), the results showed that the Belgian and Dutch lakes were generally more species-rich than the Danish lakes, whereas the species number both per lake and in total was much lower in the Spanish lakes. Recently, combined with the large database and ecological modeling, increasingly publications used climatic models to project the potential distribution of freshwater species. Comte et al. (2013) provide a review and some meta-analyses of the literature reporting both observed and predicted climate-induced effects on the distribution of freshwater fish, the results highlight the fact that, in recent years, freshwater fish distributions have already been affected by contemporary climate change in ways consistent with anticipated responses under future climate change scenarios: the range of most cold-water species could be reduced or shift to higher altitude or latitude, whereas that of cool- and warm-water species could expand or contract. Buisson et al. (2010) have projected the potential impacts of global climate change on stream fish assemblages and diversity in France, the results showed that only the scarce coldwater species (e.g., brown trout

Salmo trutta fario) were predicted to experience a strong reduction in their distributional area, whereas most cool water and warm water fish species (e.g., barbel *Barbus barbus,* European chub *Leuciscus cephalus*) were predicted to colonize many newly suitable sites located in intermediate streams or upstream. Local species richness was forecasted to increase greatly, and high turnover rates indicated fundamental changes in the structure of assemblages in the future. Moreover, the climate change could result in remarkably different impacts on the structure of fish assemblages depending on their position along the upstream-downstream gradient. Recent studies also suggested that future climate change is projected to modify the taxonomic composition of freshwater fish assemblages by increasing their overall similarity (Tisseuil et al., 2012).

Global climate warm may also enhance eutrophication and its symptoms by affecting external and internal nutrient loading, evaporation rates and water levels (Jeppesen et al., 2007; 2009). Otherwise, cascading effects of changes in the fish assemblages could also cause a series of affects on the freshwater ecosystem functioning, such as the climate-induced change of fish community would alter the food-web structure, and the predator-prey dynamics may thus be uncoupled (Lazzaro, 1997). A study contained 81 shallow European lakes from Northern Sweden to Spain indicated that the ratio of fish biomass to zooplankton biomass increased southwards, whilst the zooplankton: phytoplankton biomass ratio decreased substantially (Gyllstrom et al., 2005). Consequently, the change in fish community structure leading to increased proportions of smaller-bodied individuals will also impact other lake processes, such as nutrient dynamics and mobilization.

1.2 Lake fish diversity and the effects in China

There are numerous of lakes distributed across mainland China, according to the recently survey, more than 2759 natural lakes (area>1.0 km^2) were existed in mainland China, with the total area of 91019.63km^2 (Wang and Dou, 1998). Lakes thus were consisted as one part of the most important resources in China, with ecological functions contained water supply, entrainment, fisheries, transport and so on. Fish play a key role in the trophic dynamics and ecosystem structure of lakes, fish species richness varied considerably from 0 to 200 in different lakes (Tejerina-Garro et al., 1998; Amarasinghe and Welcomme, 2002). Lakes usually connected with rivers or other water bodies, and thus with higher fish diversity (Xie and Chen, 1999). Lakes contributed most to fish biodiversity, supporting more species, contained unique species and scarce species than other freshwater ecosystems, as rivers, streams and ponds at a regional level (Williams et al., 2003). Chinese lakes, occupying significantly different natural environments, climatic factors and fish biota due to the vast area, have always been considered as the ‘centre of dispersal’ for fish species (Wu, 1964; 1977). Through the literature review and meta-analysis on the published data, we found that there were more than 400 species were lived or living in the lakes in China (Guo unpublished data). Fish diversity in Chinese lakes contributed to one of the worldwide hotspot in the area of biodiversity study.

However, during the last several decades, freshwater fishes worldwide face an array of threats from lake and habitat degradation, dam construction, river fragmentation, pollution, over-exploitation and interactions with introduced species. Such impacts are especially prevalent in densely populated countries undergoing rapid development.

Land-use and land-cover change could affect habitat availability considerably and thus leads to the alterations of fish biodiversity (He et al., 2011; Park et al., 2006). Over the last several decades, along with urbanization and economic development, impoldering around lakes appeared in China. Increased levels of impoldering have caused a considerable loss of wild habitat, lead to shrink of lake and even disappeared, deterioration of water and air quality, and loss of biodiversity (Fang et al. 2009, Li et al., 2006). Moreover, lake environmental pollution could also lead to the habitat loss, because the water quality would affect fish lives and growth. Nowadays, large number of lakes in China has suffered from eutrophication (Cui and Li, 2005), the resultant affected lakes causes symptoms such as algal blooms, heavy growth of rooted aquatic plants (macrophytes), algal mats, deoxygenation and, in some cases, unpleasant odor, which often affects most of the vital uses of the water such as water supply, recreation, fisheries, or aesthetics.

Large dam construction affect the fish diversity mostly by causing habitat loss, alter the reproductive environments of some species, and block migration routes, and thus leading to a substantial decline in biodiversity (Stanley and Doyle, 2003; Wu et al., 2004). Such as the Three Gogres Dam in China, the largest hydroelectric dam in the world will have catastrophic consequence for fish, most especially to the migrant species in the middle and lower reaches of Yangtze River, by damaging the new spawning sites already formed below and above the Gezhou Dam and completely blocking the upstream migration routes of fish (Xie, 2003). Otherwise, dam construction between lakes and rivers could lead to lake-river separated, which also considered to be one of the reasons of fish diversity decrease (Liu et al., 2010).

Fisheries in lakes began in 1950s from fishing wild resources to stock-enhanced fish resources in China, nowadays, due to the food safe issues, lake fisheries have already spreaded all over the mainland, but mostly in the eastern China. However, along with development of fisheries in the lakes, still some problems that could also have serious negative effects on fish diversity in the lake. Such as fertilization, over-exploition, introduced exotic species. Fertilization was common in some lakes to increase the growth rate of some traditional aquaculture species (e.g. *Hypophthalmichthys molitrix* and *Hypophthalmichthys nobilis*), which can aggravate entrophication and lead to environmental pollution; Over-exploitation of fish resources appeared in lots of lakes, certain fishing techniques, such as dense-aperture nets, bombing, poisoning, and electric fishing, has severely affected breeding and regeneration of fish species in lakes (Cui and Li, 2005). Large scale of artificial stocking of introduced fish species would alter the fish richness and diversity without doubt; however invasion of exotic species is an important causal factor in the threats to some native or endemic species, because invaders usually have a fast growth rate, strong resistance to extreme environments, a wide food spectrum, and a high reproductive rate. The impact of introduced species as predators or competitors of native species has been especially striking in many plateau lakes (Xie and Chen, 1999), and in many cases these exotic species could replace the native species.

1.3 Development and application of model techniques

1.3.1 Species distribution modelling

The fascinating question of how species are distributed on earth in space and time has a long history which has inspired many biogeographers and ecologists to seek explanations. Recently, increasing interest in how species respond to global changes makes it essential to be able to determine species distribution ranges accurately. However, for a long time, it is too difficult to assess and monitor organisms in a large scale, while in this case, species distribution models (SDM) which could integrate fragmental information of species and habitats globally showed it’s advantages (Guisan and Zimmermann, 2000). Species distribution modelling (sometimes called environmental or bioclimatic niche modelling) relies on ecological theory of processes that mediate species distributions and abundance-especially niche theory (Austin, 2002). Species distribution models (SDMs) are generally developed to quantify the association between species’ occurrence or abundance and environmental, climatic or geographical predictors (Guisan and Zimmerman, 2000; Araújo and Guisan, 2006; Elith et al., 2007), and different kinds of SDMs have been widely used to predict species distribution not only for conservation and management purposes, but also for forecasting the effects of environmental or climate change (Guisan and Thuiller, 2005; Heikkinen et al., 2006; Pompe et al., 2008; Elith and Leathwick, 2009; Kharouba et al., 2009). Over the last decade, numerous of SDMs as Generalized Linear Models (GLMs) (McCullagh and Nelder, 1989), Generalized Additive Models (GAMs) (Hastie and Tibshirani, 1990), Classification and Regression Tree (CART) (Breiman et al., 1984), Random Forest (RF) (Breiman,2001), and even some multi-species model as Multivariate Adaptive Regression Splines (MARS, Friedman, 1991), Artificial Neural Networks (ANN) (Ripley, 1996) and Multivariate Regression Trees (MRT) (De’ath, 2002) have been documented and applied routinely for (1) predicting distribution patterns or assemblages from the current habitat status (Park et al., 2006; Buisson et al., 2008; Grenouillet et al., 2011), (2) evaluating the potential spreading capacity of invasive species (Roura-Pascual et al., 2009; Brummer et al., 2013), and (3) assessing biological responses and occurrences to global changes (Thuiller et al., 2005; Araujo et al., 2006; Buisson and Grenouillet, 2009; Tisseuil et al., 2012; Yu et al., 2013). With the rapid development of computing capacity and large databases, the SDMs have been widely used in both basic and applied ecology (Fig.1).

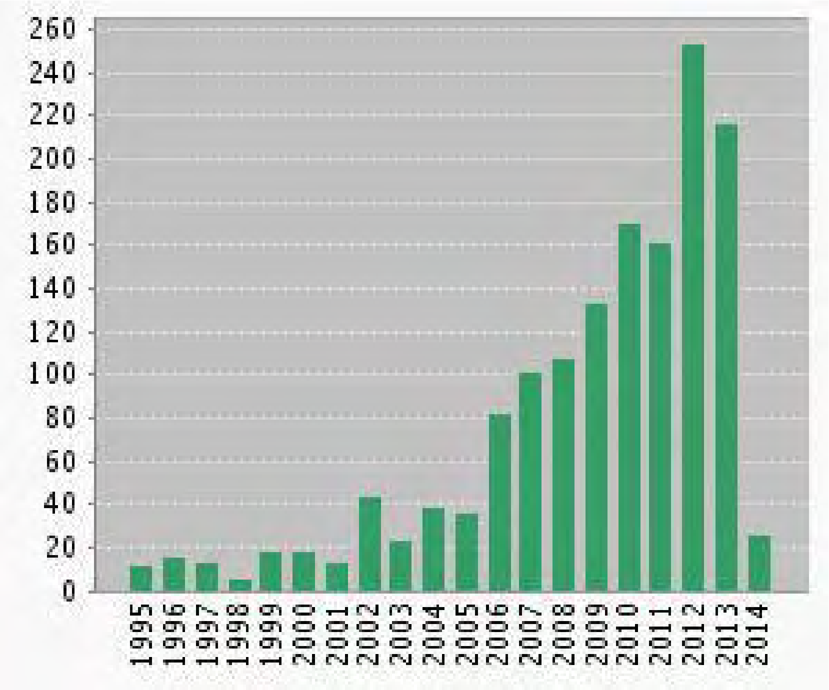


Fig.4 Publications focus on species distribution models. Data source from Web of Knowledge

To build a species distribution modelling, the principal steps are outlined in Figure 5. Normally two types of input data are needed: 1) Biological data; and 2) Environmental variables.

Biological data: Species distribution data could be either presence-absence data (records of presence and absence of the species at sampled localities) or presence-only data (records of localities where the species has been observed). Normally, species distribution data may be obtained in a variety of ways (Richard, 2007):

1) *Personal collection*: occurrence localities can be obtained during field surveys by an individual or small group of researchers.

2) *Museum collections*: occurrence localities can be obtained from collections in natural history museums.
3) *Online resources*: distribution data from a variety of sources are increasing being made available over the internet.
4) *Published data*: occurrence data from a published database

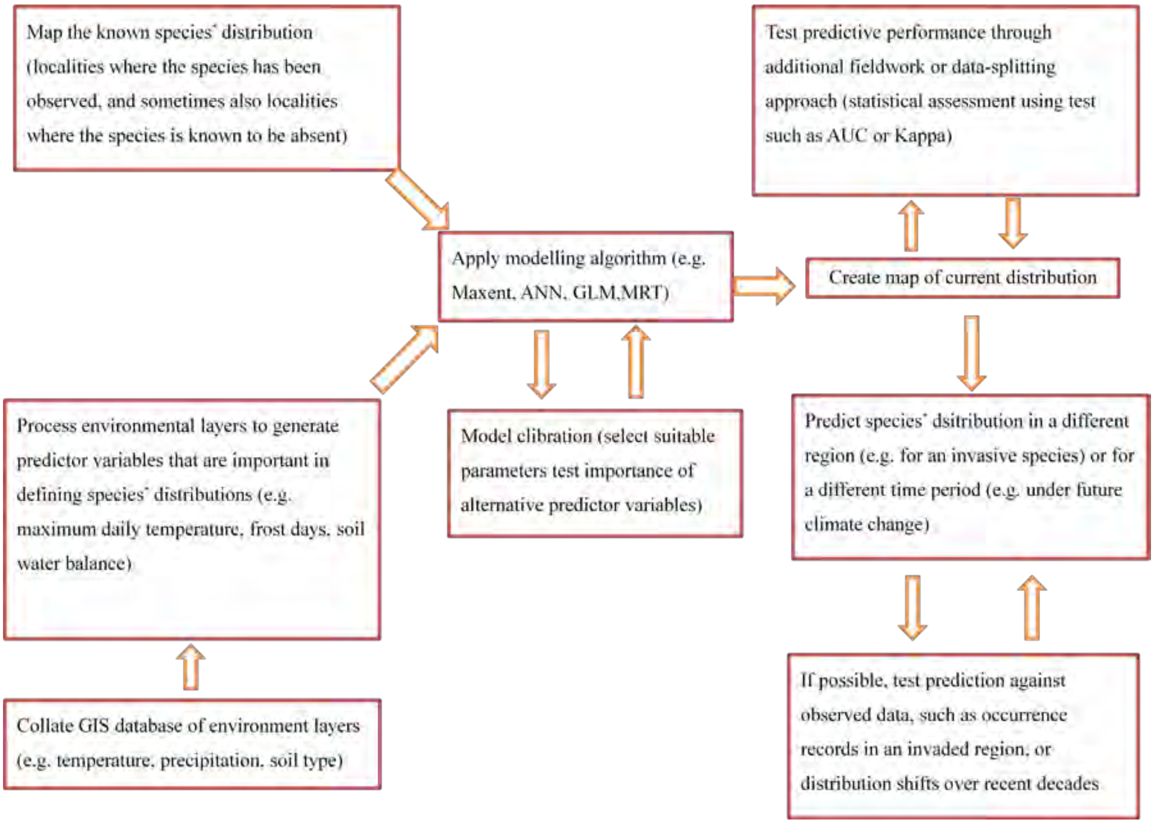


Fig.5 Flow diagram detailing the main steps required for building and validating a correlative species distribution model, Modeified from (Richard, 2007)

Environmental data

A wide range of environmental input variables have been employed in species' distribution modeling. Most common are variables relating to climate (e.g. temperature, precipitation), topography (e.g., elevation, aspect), soil type and land cover type. Variables tend to describe primarily the abiotic environment, although there is potential to include biotic interactions within the modeling. Modern technologies, including RS (remote sensing), the internet, and GIS have greatly facilitated the collection and dissemination of environmental datasets; In addition, global climate models have been used to generate scenarios of future climates and to simulate climatic conditions since the end of the last glacial period.

1.3.2 Ecopath with Ecosim

Ecopath with Ecosim, also named as static mass-balance trophic model (freely available at www.ecopath.org; Christensen et al., 2004), which focuses on energy transfer between trophic levels and is now widely used in aquatic ecosystem research (Fetahi et al., 2011; Li et al., 2009; Byron et al., 2011; Xu et al., 2011a; Xu et al., 2011b). It differs from other modeling approaches because it encompasses the full trophic spectrum, which makes it appropriate for quantitative assessment of ecosystem structure and function systematically (Christensen, 1995). The basic mass-balance equation of EwE can be described as:

Production = catches + predation mortality + biomass accumulation + net migration + other mortality or re-expressed concisely and intelligibly as:

$$B_i \cdot \left(\frac{P}{B}\right)_i \cdot EE_i - \sum_{j=1}^{n} B_j \cdot \left(\frac{Q}{B}\right)_j \cdot DC_{ji} - EX_i = 0 \qquad (1)$$

where B_i is the biomass of group i; $\left(\frac{P}{B}\right)_i$ represents the Production/Biomass ratio of group i, which is equal to the coefficient of total mortality Z under steady-state conditions (Allen 1971); EE_i is the ecotrophic efficiency of group i; B_j is the biomass of predator j; $\left(\frac{Q}{B}\right)_j$ is the consumption/biomass ratio of predator j; DC_{ji} is the contribution of prey i in the diet of predator j; EX_i is the export of group i (Christensen et al., 2004; Christensen et al., 2005). For each functional group, the composition of the diet and at least three of the four parameters (B, P/B, EE, and Q/B) must be inputed to establish the model, and the other parameters which are unknown can be estimated by the model.

Ecopath with Ecosim (EwE) has been widely considered to be an appropriate tool for the analysis of food webs. The Ecopath model is ecosystem-based software, and like other ecosystem models, can summarize and view changes in species interactions within an ecosystem (Christensen et al., 2005; Xu et al., 2011b). Both direct and indirect effects of species changes in the ecosystem can be explored and effects on the overall functioning of the ecosystem can be estimated (Heymans et al., 2004). Furthermore, the suitability of this approach is its application to a broad field of theories that are useful for ecosystem studies, e.g. thermodynamic concepts, information theory, trophic level description and network analysis (Xu et al., 2011a). EwE was first introduced to China by Tong (1999) and thereafter has been used in several aquatic ecosystems to manage fisheries in China, especially in ocean systems. However, few EwE models have been constructed for China's lake ecosystems except Li et al. (2009) and Jia et al. (2012). So far, no EwE model is available for the shallow macrophytic lakes in the MYRB, which have long been used for stocking fish. These lakes have unique aquatic living resources and environmental conditions.

1.4 Specific objectives

During the last several decades, in the context of global change, climate change and anthoropogenic activities have long been acknowledged as the two main determaints which drive the fish diversity and distributions patterns, and ultimately affect the acquatic ecosystem properties and structure. However, very few efforts aimed at the fish diversity and ecosystem in the lakes across China. While knowledge of the biodiversity of lake ecosystems is very incomplete, but declines in biodiversity are thought to be far greater in fresh water than in the most affected terrestrial ecosystems. Consequently, in the present study, we focused on the climate and anthoropogenic effects on fish diversity and distribution patterns as well as the ecosystem with the approach of several ecological modelling.

Specifically, the present study mostly aims to 1) determining the lake fish distribution and assemblage patterns as well as the drive factors; 2) predicting the lake fish species distribution and richness in the context of climate change; 3) ensemble modeling the lake fish species distribution and the uncertainties; 4) modeling lake ecosystem structure and properties under artificial activites.

2. Materials and Methods

2.1 Study area

2.1.1 Study area for Fish diversity and distribution

Lakes in China always had with complicated natural environments and large geological variations. In the present study, a total of 135 lakes (Fig.6, 21.9-48.9º N and 81.2-132.6º E) distributed across China with creditable fish presence-absence data were selected. The surface areas of these lakes ranged from 0.0006 to 2933 km^2. The detailed location and limnological characteristics of each lake are summarized in Appendix Table S1.

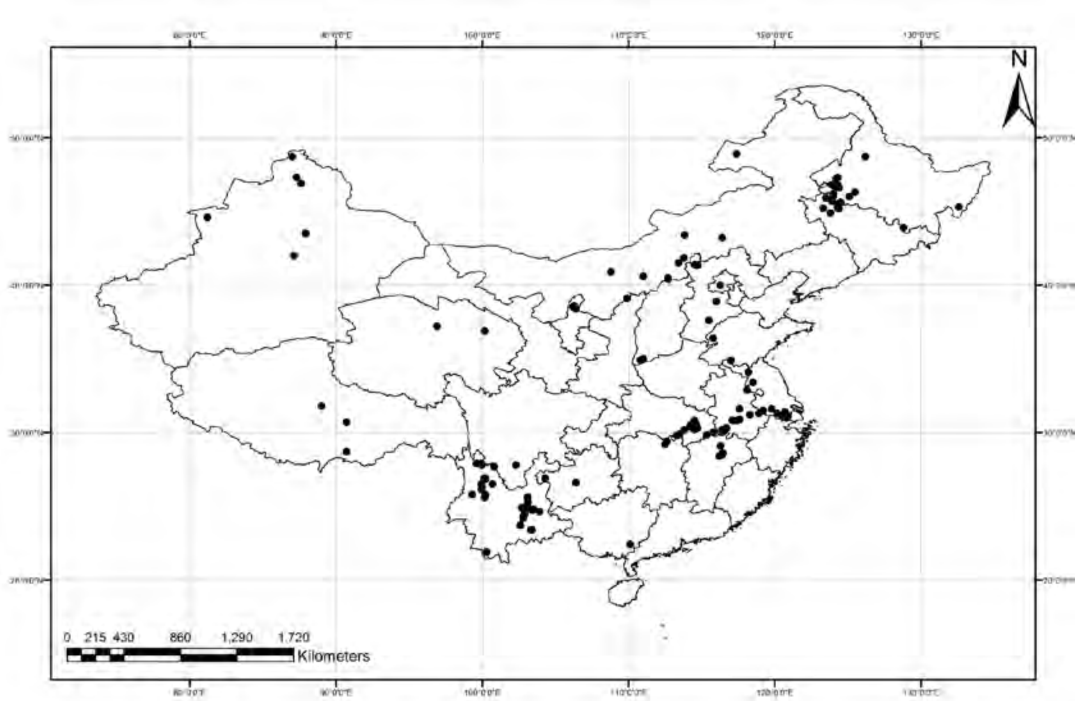


Fig.6 Map of the study lakes distributed in mainland China

2.1.2 Study area for lake ecosystem

Bao’an Lake (Fig.7), with a total area of 39.3 km^2, is a typical shallow macrophytic lake (usually with an average water depth of 2.1m) located on the south bank of the Yangtze River basin within Hubei Province, Central China (Guan, 1995). This lake used to be part of a larger lake, Liangzi Lake, which was connected to the Yangtze River by a channel. According to a survey, the area covered with aquatic macrophytes in the lake reached 75% of the whole lake, dominated by *Vallisneria spiralis* L., *Myriophyllum spicatum* L. and *Nelunbo nucefera* G. (Su et al., 1995). The physicochemical parameters of water quality during the study period are listed in Table1.

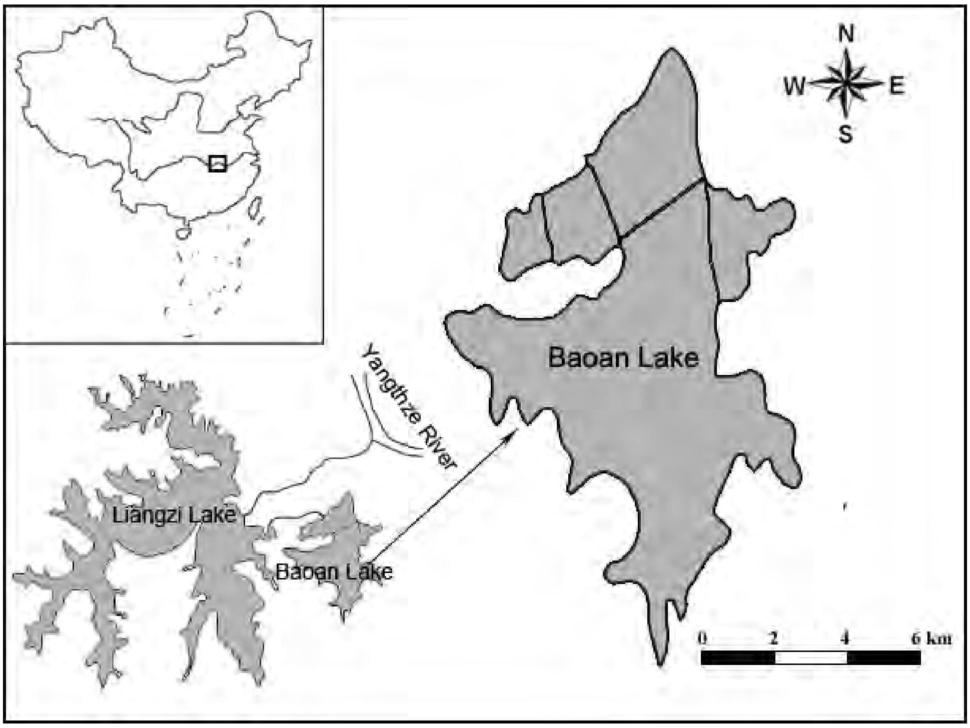


Fig.7 Geographic location and sketch map of Bao’an Lake

Tab.1 Physicochemical parameters of water quality in Bao'an Lake during 1992-1993

Parameters	Unit	Annual average
Water temperature	°C	18.9
Secchi disk depth	cm	148
pH	-	8.06
Conductivity	μs/cm	267.4
Alkalinity	-	46.54
Dissolved oxygen	mg/L	9.82
Chemical oxygen demand	mg/L	3.60
Total nitrogen	mg/L	1.460
Total phosphorus	mg/L	0.037

2.2 Data Collection

2.2.1 Fish diversity data

The presence-absence data of fish in the 135 lakes were compiled from published fish surveys since the 1950s (see Appendix S1 for details). To ensure the quality and authenticity of the data, fish compositions derived from long-term monitoring and surveying were adopted, while the data with insufficient sampling were omitted. For the lakes that had fish fauna data from several time periods, normally the earliest were preferred as they may have suffered from fewer artificial disturbances. Fish composition data were then revised according to the monograph "*Fauna Sinica*" and Wu (1977; 1964) to deal with possible nomenclatural changes (e.g. synonyms, newly discovered and named fish species). Consequently, 425 fish species and subspecies were included.

2.2.2 Climatic and environmental variables

A set of climatic and lake geographic variables were used as the explanatory variables. The climatic variables were derived from the WordClim database (Hijimans et al., 2005, available at http://www.worldclim.org) using Arcgis 10.1 (ESRI). This database was developed from compiled monthly averages of climate as measured at weather stations from a large number of global, regional, national and local sources, mostly from the 1950–2000 periods, using the Thin Plate Smoothing Spline (TPS) algorithm that yielded climate surfaces for monthly maximum, minimum, mean temperatures and total monthly precipitation (Hutchinson, 1995). In order to improve the accuracy of our analyses, we choose the highest resolution (30 arc-seconds (~1 km)). Totally 23 variables were listed and used for the fish patterns analysis. In the prediction study, a PCA (principal component analysis) was performed to eliminate the variables with high correlation, and finally 8 variables are included (Table 1, in bold).

Tab.2 The description of the predictor variables used in the study (The variables with **bold** are selected to be used in the predictive model)

Variables abbreviation	Variable type	Min	Median	Mean	Max	SD
Latitude	Mean latitude	21.90	31.10	34.42	48.90	7.90
Longitude	Mean Longitude	81.20	114.40	111.60	132.60	10.47
ALT	**Elevation of the site**	2.00	144.00	911.70	4947.00	1154.56
A	Surface area of the lake	0.01	27.95	225.80	4930.00	544.82
AMT	**Bio1:Annual mean temperature**	-5.00	140.00	113.70	217.00	61.45
MDR	**Bio2:Mean diurnal range: mean of monthly (max temp–min temp)**	70.00	110.00	104.50	146.00	19.95
ISO	**Bio3:Isothermality: (Bio2/Bio7) *100**	21.00	25.00	30.16	53.00	9.78
TS	Bio4:Temperature seasonality (SD *100)	3242.00	8807.00	9305.00	15770.00	3468.65
MTWM	**Bio5:Maximum temperature of warmest month**	130.00	286.00	284.40	340.00	46.71
MTCM	Bio6:Minimum temperature of coldest month	-295.00	-22.00	-79.59	93.00	109.16
TAR	**Bio7:Temperature annual range (Bio5–Bio6)**	212.00	331.00	364.00	561.00	103.52
MTWE	Bio8:Mean temperature of wettest quarter	78.00	216.00	214.30	283.00	40.57
MTDQ	Bio9:Mean temperature of driest quarter	-208.00	50.00	-2.47	162.00	104.59

MTWA	Bio10:Mean temperature of warmest quarter	78.00	221.00	227.00	286.00	48.12
MTCQ	Bio11:Mean temperature of coldest quarter	-213.00	43.00	-12.32	149.00	99.40
AP	Bio12:Annual precipitation	82.00	919.00	806.30	1480.00	409.61
PWM	Bio13:Precipitation of wettest month	18.00	186.00	168.60	305.00	63.03
PDM	**Bio14:Precipitation of driest month**	0.00	9.00	15.04	44.00	14.62
PS	Bio15:Precipitation seasonality(coefficient of variation)	39.00	86.00	83.35	138.00	26.64
PWEQ	Bio16:Precipitation of wettest quarter	47.00	461.00	423.70	815.00	171.80
PDQ	Bio17:Precipitation of driest quarter	2.00	38.00	56.64	160.00	55.12
PWAQ	Bio18:Precipitation of warmest quarter	47.00	443.00	403.20	676.00	157.02
PCQ	**Bio19:Precipitation of coldest quarter**	2.00	39.00	60.01	175.00	59.59

The unit for all the temperature was (°C * 10), for all the precipitation was (mm).

Min: the smallest observation**;**

Median: the median value of all the observations;

Mean: the mean value of all the observations;

Max: the largest observation;

SD: the standard deviation.

2.2.3 Data for ecosystem mass-balance model

A total of 23 functional groups were defined to establish the mass-balance model of Bao’an Lake. Groups were classified according to their trophic habits (mainly diet), abundance and the availability of information (Christensen et al., 2004). Some commercial fish species and Chinese mitten crab (*Eriocheir sinensis*) were grouped separately due to their importance to fishery yield and stocking (see details in Table 3).

For all functional groups, the key input data used to construct the model usually included biomass (B_i), production/biomass ratio (P_i/B_i), consumption/biomass ratio (Q_i/B_i), ecotrophic efficiency (EE_i) and food composition (DC_{ji}) (Christensen et al., 2005). Normally, fishery capture of some commercial products (such as fish, crabs, shrimps, etc) was also contained in the model in order to discuss the effects of fishing. Here the fishing yield data were also compiled with the monographs. Because many ecological and fisheries studies had been carried out in Bao’an Lake since the 1980s, especially during the period of 1992-1993, abundant first-hand data on the lake could be collected and compiled from published papers (Hu and Huang, 1991; Liang and Liu, 1995), dissertations (Zhang, 2005; Yan, 1998; Jin, 2003), and unpublished data.

The specific approaches to these parameters are summarized as follows.

Fish

The biomass and P/B of the small fish group in the lake were estimated directly by Zhang (1999; 2005). For other fish groups, the biomass and P/B values were calculated from the following equations:

$$B=Y/F \tag{2}$$

$$F=Z-M \tag{3}$$

$$Z=P/B=K\cdot\frac{L_\infty-\overline{L}}{\overline{L}-L'} \tag{4}$$

where B, Y, F, Z and M represent biomass, yield, fishing mortality, total mortality and natural mortality of fish; L_∞, $\overline{L}$ and L' represent the asymptotic length, mean length and cut-off length of fish (Beverton and Holt, 1957; Allen, 1971).

Tab. 3 Species composition of each functional group for the Bao’an Lake ecosystem model

NO.	Functional group	Dominant species composition
1	Mandarin fish	*Siniperca chuatsi*, *Siniperca kneri*
2	Snakehead fish	*Channa argus*
3	Large culters	*Culter alburnus*, *Culter mongolicus*, *Culter dabryi*
4	Catfish	*Pelteobagrus fulvidraco*
5	Common carp	*Cyprinus carpio*
6	Crucian carp	*Carassius auratus*
7	Small fishes	*Pseudorasbora parva*, *Sqalidus nitens*, *Sarcocheilichthys nigripinnis*, *Hemiculter leucisculus*, *Toxabramis swinhonis*, *Gobiidae*
8	Silver carp	*Hypophthalmichthys molitrix*
9	Bighead carp	*Aristichthy nobilis*
10	Grass carp	*Ctenopharyngodon idellus*
11	Breams	*Megalobrama amblycephala*, *Parabramis pekinensis*
12	Crabs	*Eriocheir sinensis*
13	Shrimps	*Caridina nilotic gracilipes*, *Macrobrachium nipponense*, *Neocaridina denticulate sinensis*
14	Molluscs	*Bellamya aeruginosa*, *Alocinma longicornis*, *P. striatulus*
15	Oligochaeta	*Branchiura Sowerbyi*, *Limnodrilus hoffmeisteri*
16	Aquatic insecta	Chironmoidae
17	Microzooplankton	Rorifers, Protozoans
18	Cladocera	*Diaphanosoma leuchtenbergianum*, *Sida crystalline*, *Moina micrura*
19	Copepoda	*Cyclops vicinus*, *Mesocyclops leeckarti*
20	Submerged macrophyte	*Potamogeton maackianus*, *Myriophyllum spicatum*, *Vallisneria spiralis*, *Ceratophyllum demersum*
21	Phytoplankton	Chlorophyta, Bacillariophyta, Cyanophyta
22	Attached algae	Chlorophyta, Bacillariophyta, Cyanophyta
23	Detritus	Bioseston, Abioseston

The consumption/biomass ratio (Q/B) of fish was estimated from the empirical equation presented by Palomares and Pauly (1998):

$$\log(Q/B) = 7.964 - 0.204 \log W_{\infty} - 1.965 T' + 0.083 A + 0.532 h + 0.398 d \qquad (5)$$

where W_{∞} is the asymptotic weight (g), T' the mean temperature of the lake expressed by T'=1000/Kelvin (Kelvin=°C+273.15), A is the aspect ratio ($A=H^2/S$, H is the height of caudal fin and S is the surface area) for a given fish, h is a dummy variable expressing food type (1 for herbivores, and 0 for detritivores and carnivores), and d is a dummy variable also expressing food type (1 for detritivores, and 0 for herbivores and carnivores). The morphometric data for estimating Q/B was derived from Zhang (2005). The diet compositions (DC_{ji}) of fish groups were also modified from Zhang (2005).

Macrocrustaceans

The biomass of the Chinese mitten crab was calculated according to its stocking biomass and recapture rate surveyed during 1991-1993 in the lake (Liu et al., 1995; Jin, 2001), and the biomass of shrimps was estimated directly by Qin et al. (2003) with a pop-net. Since there has been no exact P/B or Q/B data for the macrocrustaceans in the lake until now, we assumed the value of EE and P/Q to be 0.85 and 0.075 respectively according to a similar study in Lake Taihu (Li et al., 2009). The dietary composition came from Jin et al. (2001) for Chinese mitten crab and Li et al. (2009) for shrimps.

Zoobenthos and Zooplanktons

An extensive and detailed plankton survey had been carried out during 1992-1993 in Bao’an Lake and the relevant results were compiled in the book “Resource, environment and fishery ecological management of macrophytic lakes” (Liang and Liu, 1995). Therefore, the biomasses of molluscs, oligochaetes, aquatic insecta, microzooplankton, cladocera, copepoda were taken directly from the published materials. The P/B ratio of mollucks, oligochaetes and aquatic insecta in the lake were estimated directly by Yan et al. (1998), and the P/Q were assumed to be 0.05 (Yan et al. 2003). For the groups of microzooplankton, cladocera, copepoda, 0.05 and 0.95 were adopted for the values of P/Q and EE (Park et al., 1974; Scavia et al., 1974). All the dietary compositions were modified from Liu (1999) with slight adjustments to balance the model if necessary.

Primary producers and detritus

Phytoplankton, attached algae and submerged macrophytes were defined as the three groups of primary producers. The biomasses of the three groups were investigated directly and reported in the published book (Liang and Liu, 1995). A P/B ratio of 185 year^{-1} for phytoplankton (Sun et al., 1999), 250.7 year^{-1} for attached algae (Wang et al., 1995) and 1.25 year^{-1} for macrophytes (Liu, 1992) were used in the same lake. The biomass of detritus was also calculated directly in the lake by Hu and Huang (1991).

2.3 Model Techniques

2.3.1 Species distribution mdoelling

In our study, a novel technique for modeling species-environment relationships, MRT, was used as the main tree model to explore the relationships between fish species and environments. MRT is the tree model based on the same recursive partitioning principles as Classification and Regression Trees (CART) but extended to multivariate response variables, which is an important consideration for the community studies. MRT splits objects (e.g. sampling sites) into homogenous groups according to the response, with the splits constrained by explanatory variables. The tree is grown by splitting the data a large number of times, and then it is subsequently pruned (reduction of the number of groups) via a re-sampling method called v-fold cross-validation (Breiman et al., 1984) to obtain the best predictive tree size. The model was run with 100 replicates to be certain that the results were not simply obtained by chance. An unconstrained cluster (hclust) was also presented to compare with the MRT groups, a Kruskall-Wallis test and multiple comparison tests were then conducted in order to assess the differences in species richness between each assemblage (Giraudoux, 2006). Here, the new R package “MVPART wrap” (Ouellette et al., 2012) was used instead of the package “mvpart” (De’ath, 2002) to get more detailed information from the MRT model.

Moreover, in the context of ensemble modelling, a total of nine different statistical models, contained: Generalized Linear Models (GLM, McCullagh and Nelder, 1989), Generalized Additive Models (GAM, Hastie and Tibshirani, 1990), Classification and Regression Tree (CART, Breiman et al., 1984), Random Forest (RF, Breiman, 2001), Multivariate Adaptive Regression Splines (MARS, Friedman, 1991), Artificial Neural Networks (ANN, Ripley, 1996), RF, Mixture Discriminant Analysis (MDA), Generalized Boosted Models (GBM), SRE (Surface Range Envelop or usually called “BIOCLIM”, Busby 1991), were used to predict the presence-absence data of each fish species in the assemblage. All the models were performed in R (R Development Core Team 2013) with the package of “biomod2” (Thuiller et al., 2003). For each of the 92 species, nine models were constructed respectively with a random 70% subset, and the remaining 30% dataset were used to evaluate the model performance. The split-sample procedure was repeated 100 times for each species (Fig.8).

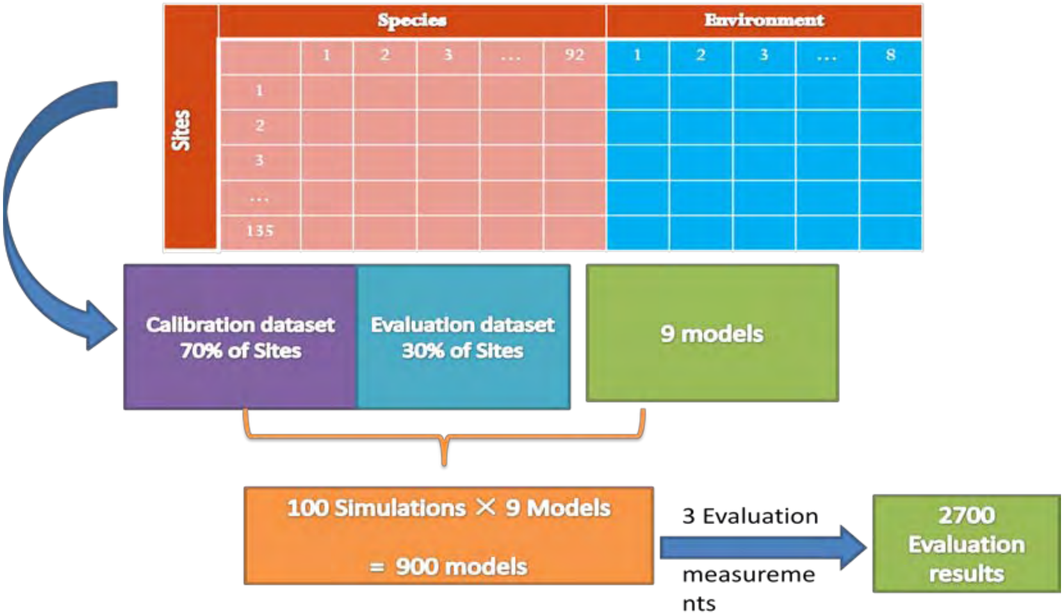


Fig.8 General schematic representation of the ensemble modelling approach

2.3.2 Ecopath modelling approach

A static mass-balance trophic model was constructed using the Ecopath with Ecosim, version 6.2 (freely available at www.ecopath.org; Christensen et al., 2004).

2.3.3 Model evaluation

Evaluation for Species distribution modelling

For the species distribution modelling, the predictive model performance were evaluated using three main kinds of accuracy measures: Area under the receiver operating characteristic curve (AUC; Fielding and Bell, 1997), Cohen’s Kappa (Cohen, 1960) and True skill statistic (TSS; Bohning et al., 2008).

Area under the receiver operating characteristic curve (AUC) is one kind of threshold-independent measures, A ROC (receiver operating characteristic curve) plot is obtained by plotting all sensitivity values (true positive fraction) on the y axis against their equivalent (1-specificity) values (false positive fraction) for all available thresholds on the x axis. The area under the ROC function (AUC) is usually taken to be an important index because it provides a single measure of overall accuracy that is not dependent upon a particular threshold (Deleo, 1993; Fielding and Bell, 1997). The value of the AUC ranged from 0 to 1, with a score of 1 indicating a perfect fit, scores from 0.9 to 1 representing an excellent fit, scores from 0.8 to 0.9 representing good fit, and those scores from 0.5 to 0.6 implying discrimination that was no better than random (Swets, 1988).

Cohen’s kappa (1960) has been adopted to alleviate the problem of overestimating accuracy. It measures the extent to which the agreement between observed and predicted is higher than that expected by chance alone. This statistic is used to assess inter-rater reliability when observing or otherwise coding qualitative / categorical variables. Kappa is considered to be an improvement over using % agreement to evaluate this type of reliability. Kappa has a range from 0-1.00, with larger values indicating better reliability.

TSS (True skill statistic) is defined as the average of the net prediction success rate for presence sites and that for absence sites. It has gained considerable theoretical interest over many years (Bohning et al., 2008), and it is considered the best available summary measure of model performance in medical diagnostic tests by some researchers (Biggerstaff, 2000). TSS takes into account both omission and commission errors, and success as a result of random guessing, and the values range from −1 to 1, where 1 indicates perfect agreement and values of zero or less indicate a performance no better than random. This index is closely related to the arithmetic mean of sensitivity and specificity.

Overall accuracy (OA), defined as the probability that a site (either presence or absence) is correctly predicted, is the most common measure used in various disciplines including ecology.

Evaluation for the Ecopath Model

After all the necessary parameters had been entered, the model could be balanced by checking the estimated values: are the EE values possible (less than 1) and are the GE (=P/Q) values physiologically realistic (0.1-0.3) for most groups? Usually, in order to balance a model we modified the input data manually: subtle adjustment should always be made, especially for diet composition (Christensen et al., 2005).

To facilitate this task and to make the process more transparent, we implemented a ‘pedigree’ routine (Funtowicz and Ravetz, 1990), which serves a dual purpose by describing the origin of the data and by assigning confidence intervals to data based on their origin (Pauly et al., 2000). This index can be understood as a coded statement categorizing the origin of a given input and specifying the likely uncertainty associated with it (Christensen et al., 2005). For each input value, a description should be made on the data source and its confidence (high or low precision, estimated by model, "guesstimated", approximate or indirect method, from other models, or from literature, etc.). Percentage ranges of uncertainty, based on a set of qualitative choices relative to the origin of B, P/B, Q/B, catch and diet composition input were used in the routine and resulted in an index value scaled from 0 (not rooted in local data) to 1 (fully rooted in local data) for each input data point. Based on the individual pedigree index values, an overall ‘pedigree index’ (P) of the information in ECOPATH was calculated:

$$P = \sum_{i=1}^{n} \sum_{j=1} \frac{I_{ij}}{n} \tag{6}$$

where I_{ij} is the pedigree index for model group i and parameter j, n is the total number of modeled groups (Christensen and Walters, 2004).

Apart from the pedigree index (P), we also used a measure of fit (t*) to scale the model uncertainty based on the number of living groups in the ecosystem. The measure of fit describes how well a given model is rooted in local data, and is calculated as:

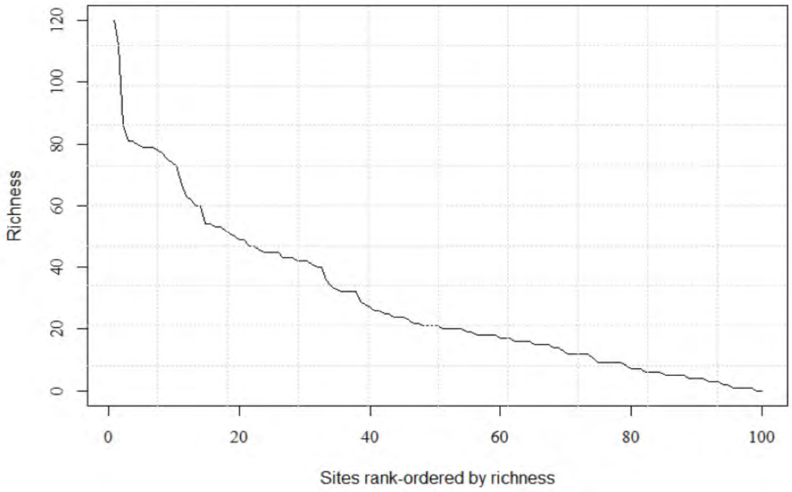
$$t^* = P * \sqrt{(n-2)} / \sqrt{(1-P^2)}$$

3. Results

3.1 Large-scale patterns of fish species distribution and assemblage

3.1.1 Comprehensive description of fish composition

A total of 425 species and subspecies from 15 orders, 42 families had been recorded in the 135 studied lakes. Rank-order species richness is sigmoid (Fig.9a): around 5% of the sites had high species richness (species richness >80), 45% contained moderate richness (20<species richness < 80), and 50% with low richness (species richness< 20). While Fig.9b shows that most of the recorded species are uncommon or rare species, about 80% of the species have an occurrence of less than 20% in all the sites, among which around 47% of the species occurred only once. About 8% of the species are recorded as moderately common species that occurred in more than 30% of the sites. Only 8 species occurred in more than 80 sites and are thus recorded as the most abundant species in our research (*Pseudorasbora parva* (Temminck et Schegel, *Hypophthalmichthys molitrix* (Cuvier et Valenciennes), *Ctenophaxyngodon idellus* (Cuvier et Valenciennes), *Aristichthys nobilis* (Richardson), *Carassius auratus* Linnaeus, *Cyprinus(cyprinus) carpio* Linnaeus, *Misgurnus anguillicaudatus* (Cantor), *Hemiculter leucisculus* (Basilewsky)).



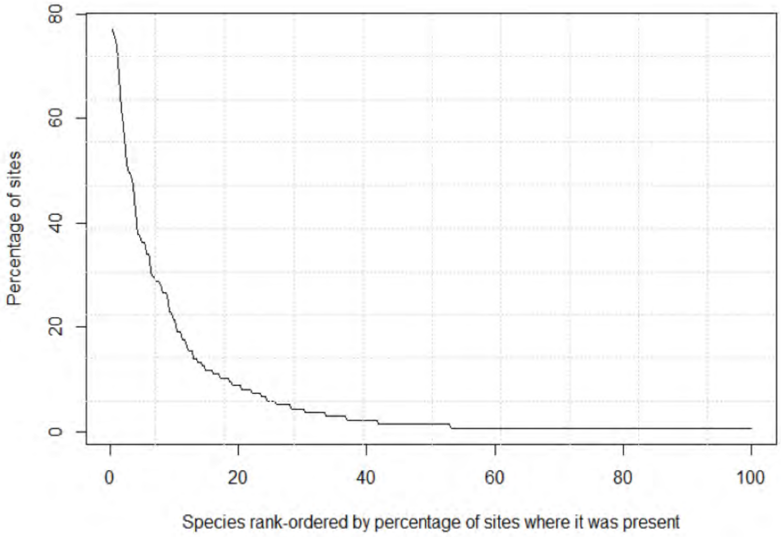


Fig. 9 Rank-ordered distribution plots of fishes in Chinese lakes

a Sites rank-ordered by species richness in each site (X axis represent the % of the 135 sites; Y axis represent the richness per site);

b Species (percentage) rank-ordered by percentage of sites where each species encountered (X axis represent the % of 435 species; Y axis represents the total percentage of sites where the species present)

3.1.2 Fish diversity and assemblages in Chinese lakes

In the MRT model, a robust decision tree with five terminal nodes was identified to minimize the cross-validated relative error and the complexity instead of the "1-se rule" as suggested by Breiman *et al.,* (1984) (Fig. 10a). Comparisons of the constrained clustering from MRT models with similar numbers of unconstrained cluster groups (using K-means clustering and Euclidean distance) showed that the constrained and unconstrained groups are similarly homogeneous, which indicates the species clusters were spatially contiguous and the environmental variables adequately accounted for the species variance (De'ath, 2002). Although the MRT tree only accounted for 35.2% of the variation in the species data (Fig. 10b), De'ath described it as common because of the large number of low occurring species (Devantier *et al.*, 2006).

Therefore, all the 135 lakes were patterned into five assemblages according to the similarity of fish composition and the determinants (Fig. 10b). Two main clusters were defined as the first split (Cluster **I** and Cluster **II**), and this split identified two regions: high altitude region (plateau) and low altitude region (plain). In addition, five sub-clusters which we considered as five assemblages were also sub-divided, and contained 61, 26, 38, 4, 6 lakes respectively, named **I**a, **I**b, **II**a1, **II**a2, **II**b.

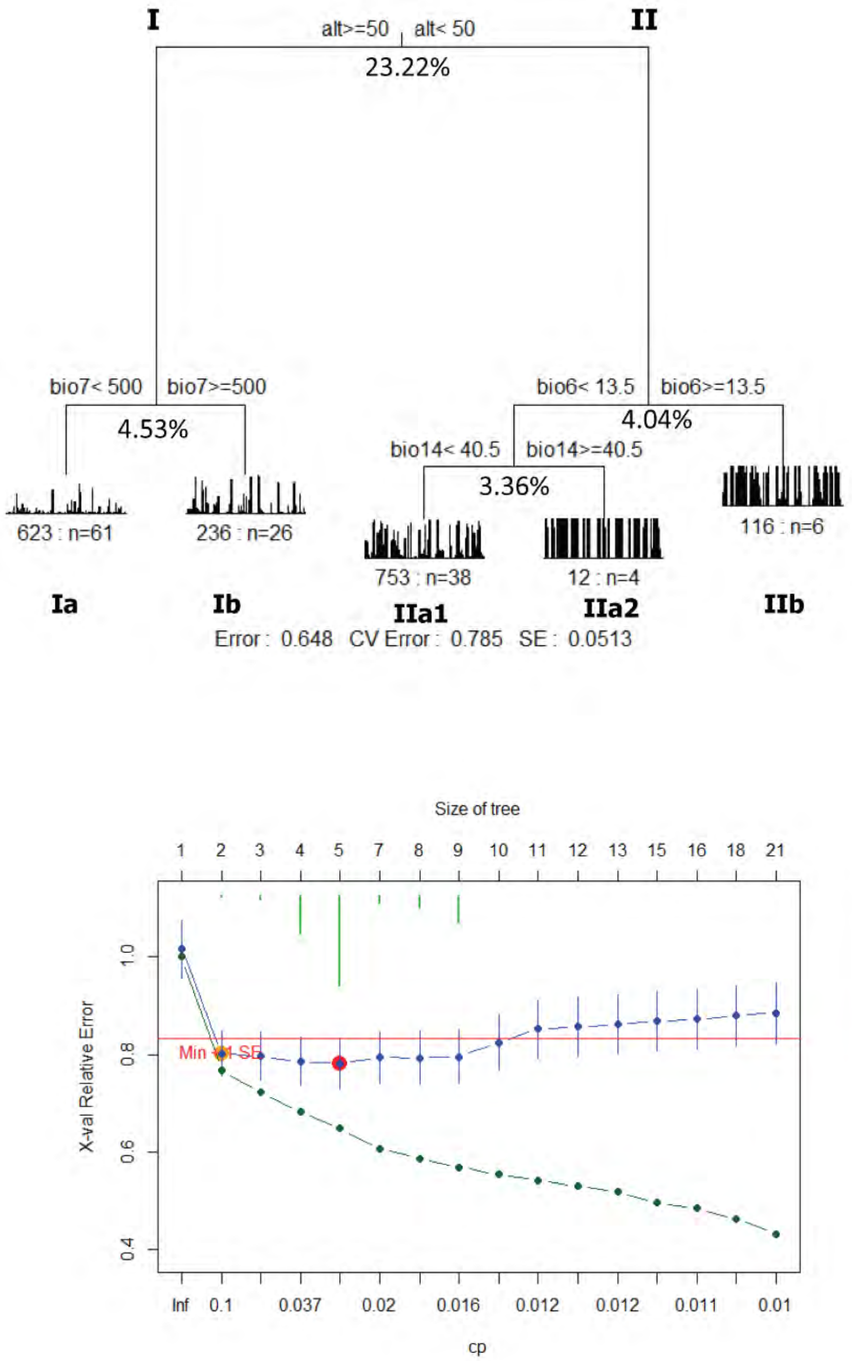


Fig. 10 The output of MRT model for the fish species and environmental variables

a Fish distribution patterns and assemblages defined by MRT model (alt: Altitude of the site; bio6: Minimum temperature of coldest month; bio7: Temperature annual range; bio14: Precipitation of driest month; the terminal nodes represent five fish assemblages **Ia, Ib, IIa1, IIa2, IIb,** see text for details)

b Cross-validation of the Multi-variable Regression Tree analysis (the upper line is cross-validation relative error while the lower line is re-substitution error, the red circled point is the model with the greatest cross-validated predictive accuracy)

Box-plots of species richness in each assemblage are shown in Figure 11. The Kruskall-Wallis test revealed that species richness varied significantly among the different assemblages (p<0.05), while the multiple comparison test showed that species richness in each of the three assemblages in cluster **II** were significantly different from the two assemblages in cluster **I** (p<0.05). However, there is no significant difference between each assemblage in the same cluster. To simplify, we defined the richness in each lake as “lake richness” while the diversity in the assemblage as “fauna diversity” hereinafter. Cumulative species richness showed that only **I**a did not reach an asymptote compared to the other four (Fig. 12). However, **I**a contained the highest rates of species’ accumulation as well as the highest fauna diversity (280 species; abbreviated as sp. hereinafter), although the species richness in each lake was the lowest (median=12 sp.; Fig.3 and Fig.4). Lake richness in **II**a2 and **II**b showed higher values (median values are 84 sp. and 88.5 sp. respectively) but with lowest accumulative rates, while **II**a2 also contained the lowest fauna diversity with only 84 species. Assemblage **II**a1 had moderate lake richness (median value = 55 sp.) as well as for the fauna diversity (184 sp.). Unfortunately, **I**b seemed to be the one that contained both lower lake richness (median value = 22.5 sp.) and fauna diversity (99sp.) than the other four assemblages (Fig.5 and Fig.6).

3.1.3 Determinants of fish distribution and assemblages in Chinese lakes

Five assemblages were thus mapped (Fig. 13) geographically: cluster **I** (**I**a and **I**b) mostly indicated the lakes distributed in Meng-Xin plateau, Yun-Gui plateau, Qinghai-Tibet plateau and the North-East plain, while cluster Ⅱ(Ⅱa1, Ⅱa2, Ⅱb) mainly indicated the lakes distributed in the Eastern plain, with the majority along the Yangtze River basin. From the tree (Fig.10b), we can see that altitude accounted for most of the variation in species distribution (23.22%), and explained about 65.9% of the total variation. Followed by the annual temperature range (bio7; accounted for 4.54%), the minimum temperature of the coldest month (bio6; accounted for 4.04%) and precipitation in the driest month (bio 14; accounted for 3.36%).

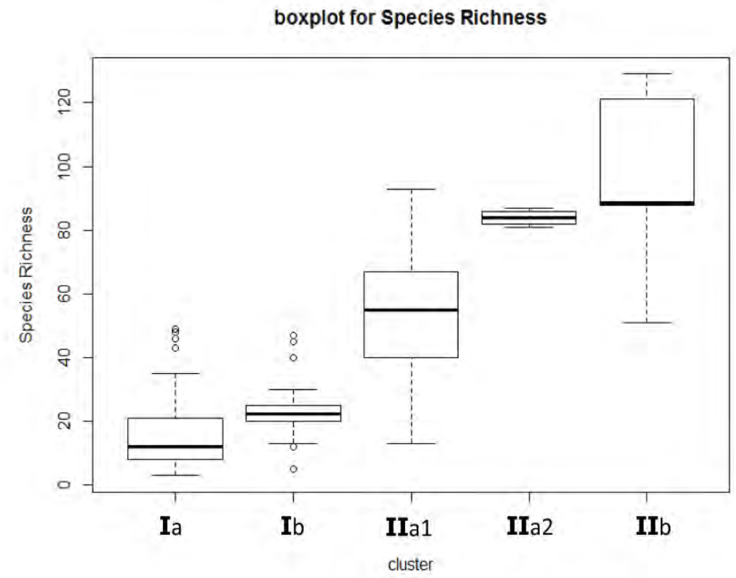


Fig. 11 Box-plot for the species richness of each assemblage (Ia, Ib, IIa1, IIa2, IIb were the five assemblages defined in Fig.4; for each box plot hereinafter in the paper, the median (line within the box), first and third quartiles (box), non-outlier range (whiskers), and outliers (dot) are shown).

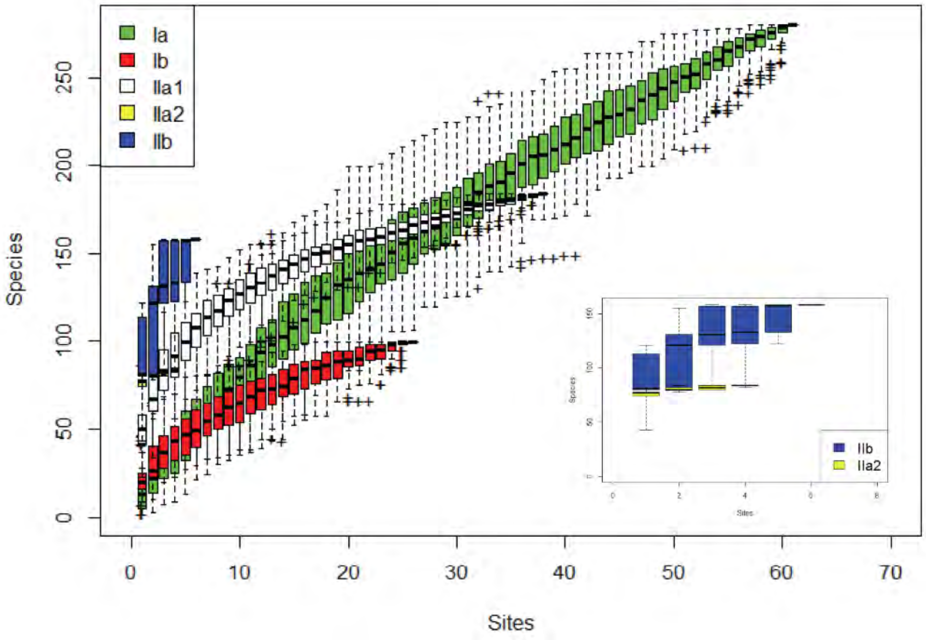


Fig. 12 Species accumulation curves for the five assemblages (figure **a** shows the global picture of the species accumulation for all the five assemblages, while the small figure **b** show the precise details of IIa2, IIb, not clearly shown in figure **a**.)

For these four determinants, **Ia** was characterized by high altitude (1908.85 ±

1050.96 m) and lower annual temperature range (32.58±10.54℃), while **Ib** in terms of relative high altitude (220.15±202.38m) and higher annual temperature range (51.54±1.47℃). However, clusterⅡ,Ⅱa1 was characterized mostly by the lower altitude (19.45±19.25m), a lower minimum temperature of the coldest month (-0.71±2.62℃) and lower precipitation in the driest month (29.97±9.28mm), while **Ⅱa2** was characterized by lower altitude (14±0.82m), lower minimum temperature in the coldest month (1.12±0.12°C) and higher precipitation in the driest month (42.75±1.5mm). Assemblage Ⅱ **a2** was divided from assemblage Ⅱ a1 mainly depending on the higher precipitation of the driest month. In Ⅱ **b**, lakes were determined in terms of lower altitude (21.16±5.84m) and higher minimum temperature of the coldest month (2.1±3.84°C).

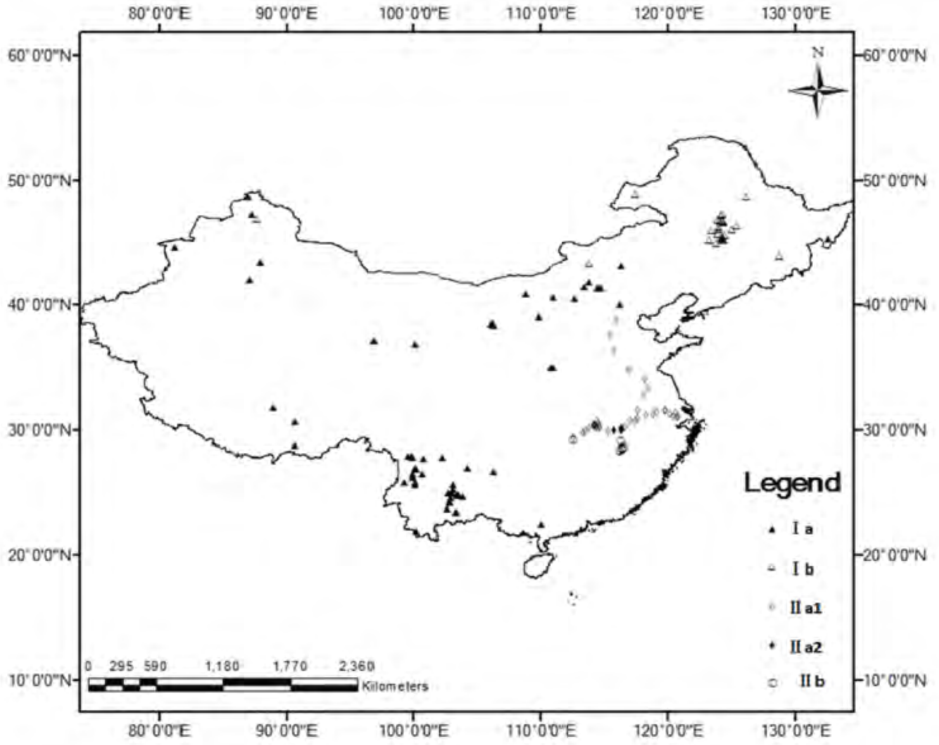


Fig. 13 Map of the five assemblages displayed in mainland China

3.1.4 Indicator species of assemblages

A total of 107 indicator species were indentified in five assemblages (contained 1 sp., 8 sp., 1 sp., 59 sp. and 38 sp. respectively) (see Table 3 for details). *Cyprinus (cyprinus) chilia* (Wu *et al.*) is the only indicator species for **Ia**, since this assemblage mostly consisted of the plateau fish fauna: fish distributed in these lakes were mostly stenochoric species, with some species even occurring only once. The main indicator species in assemblage **Ib** were *Rhodeus sericeus* (Pallas), *Sarcocheilichthys lacustris* (Dybowsky), *Ladislavia taczanowski* (Dybowsky), *Gobio lingyuanensis* (Mori), *Gobio gobio cynocephalus* (Dybowsky), *Carassius auratus gibelio* (Bloch), *Misgurnus mohoity* (Dybowski), *Perccottus glohni* (Dybowski). *Coilia nasus* (Schlegel) was the indicator species of **Ⅱa1**. Notably, cluster **Ⅱa2** and **Ⅱb** comprised only 10 of the 135 lakes but contained 92.3% of the total indicator species, with all the indicator species in these two assemblages being common species in the shallow lakes along the middle-lower Yangtze River. Therefore, **I**a contains most of the plateau species, **Ⅱa1**, **Ⅱa2** and **Ⅱb** contained mostly the plain fish, while **I**b seems to be the transition zone between plateau and plain as it contained both plain fish and plateau fish.

3.2 Predicting fish species richness and assemblages in current status

An MRT model was constructed with six terminal nodes which can be qualified as six clusters of fish assemblages (Fig. 14a). The tree size was selected by cross-validation which showed the lowest predictive error (Fig. 14b) in the calibration dataset. The coefficient of determination (R^2) is a commonly used measure of explanatory power in linear modeling which represents the proportion of variation of the response variables explained by a model. In this model R^2=47.7%, which showed that all the variables we used in the model can explain 47.7% of the total variance. Ultimately, six fish assemblages were mainly determined by four explanation variables. Overall, altitude (ALT, 30.43%) contributed the most to fish assemblages

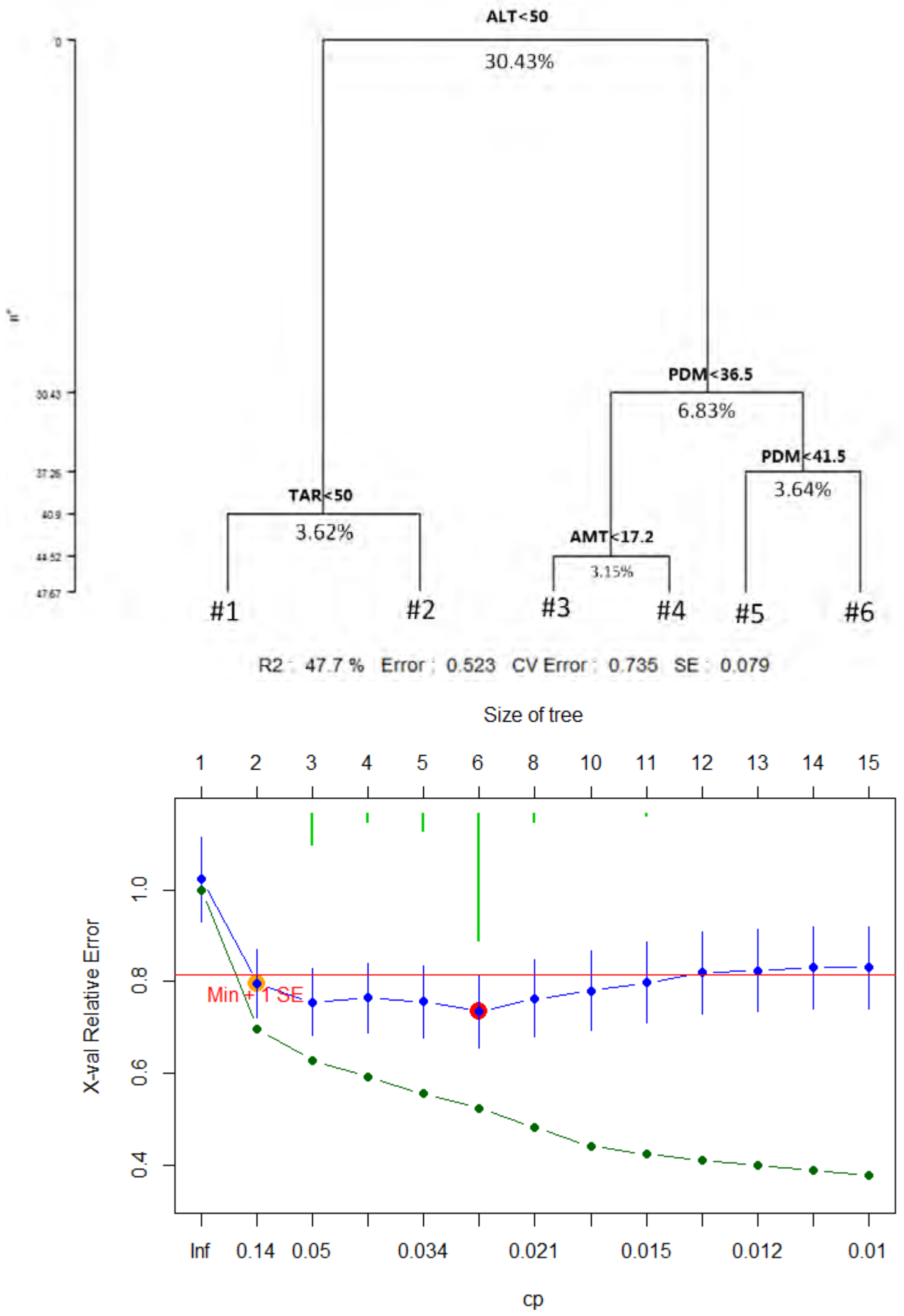


Fig.14 Multivariate response tree validated model for lake fish species distributed in China

a. MRT tree model predicting the species distributions in Chinese lakes (see Table 2 for details of the variables) **b**. Selection of the tree size by cross-validation method

Followed by precipitation of the driest month (PDM, 10.47%), annual temperature range (TAR, 3.62%) and annual mean temperature (AMT, 3.15%) (Fig.14a). Assemblage 1 and 2 contained the fish species which were linked to an altitude greater than 50 m and separated by an annual temperature range lower or higher than 50℃. Fish species in assemblage 5 and 6 were related tightly at altitudes lower than 50 m, and a high precipitation during the dry season (superior than 36.5 mm). However, fish species in assemblage 3 and 4 were determined by a relatively lower altitude, and lower precipitation during the dry season (< 36.5 mm), while assemblage 3 preferred a mean annual temperature lower than assemblage 4 (Fig. 14a).

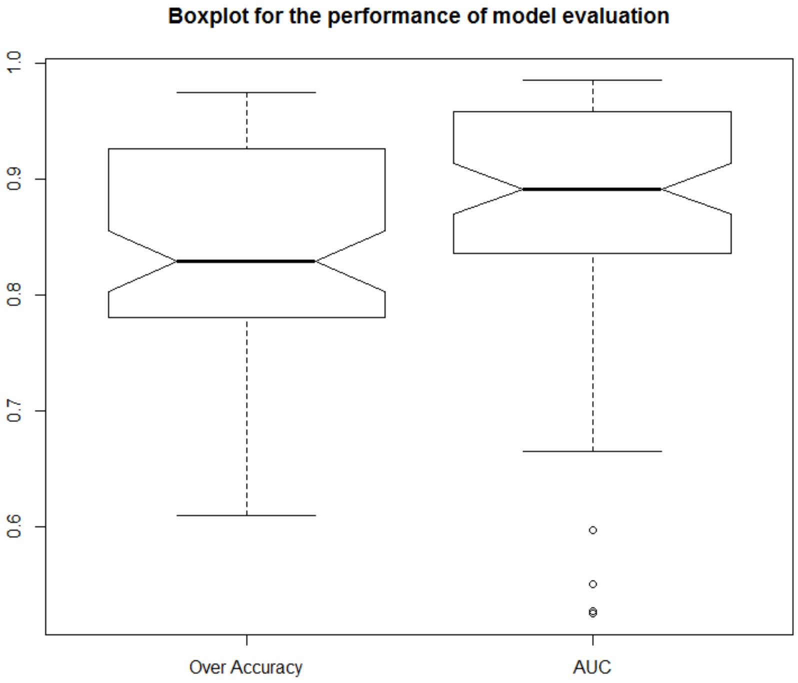


Fig.15 The evaluation results of the MRT model by Overall Accuracy and AUC measures

In order to evaluate the model performance, AUC (threshold-independent measure), overall accuracy (threshold-dependent) and Cohen's Kappa were used. The results showed that the MRT model could accurately predict the fish species distributions in terms of assemblage level, with AUC values ranging from 0.53 to 0.99 (with a median value of 0.87), the overall accuracy ranging from 0.61 to 0.98 (with a median value of 0.84) (Fig. 15), and the Kappa values range from 0.012 to 0.91 (with a median value of 0.55) (Fig. 16). All three accuracy measures showed significantly that most of the species were accurately predicted by the model. However three species (*Opsariichthys uncirostris bidens*, *Clarias fuscus* and *Paracheilognathus iimberbis*) were not accurately predicted, with kappa values lower than 0.2. This indicated the predictions were in ‘slight agreement’, while the other 74 indicator species were predicted from ‘fair agreement’ to ‘perfect agreement’ (Fig. 10). These evaluated results imply that the discrimination of the MRT model was no better than random and thus demonstrated to be robust and informative for the prediction of fish species distributions.

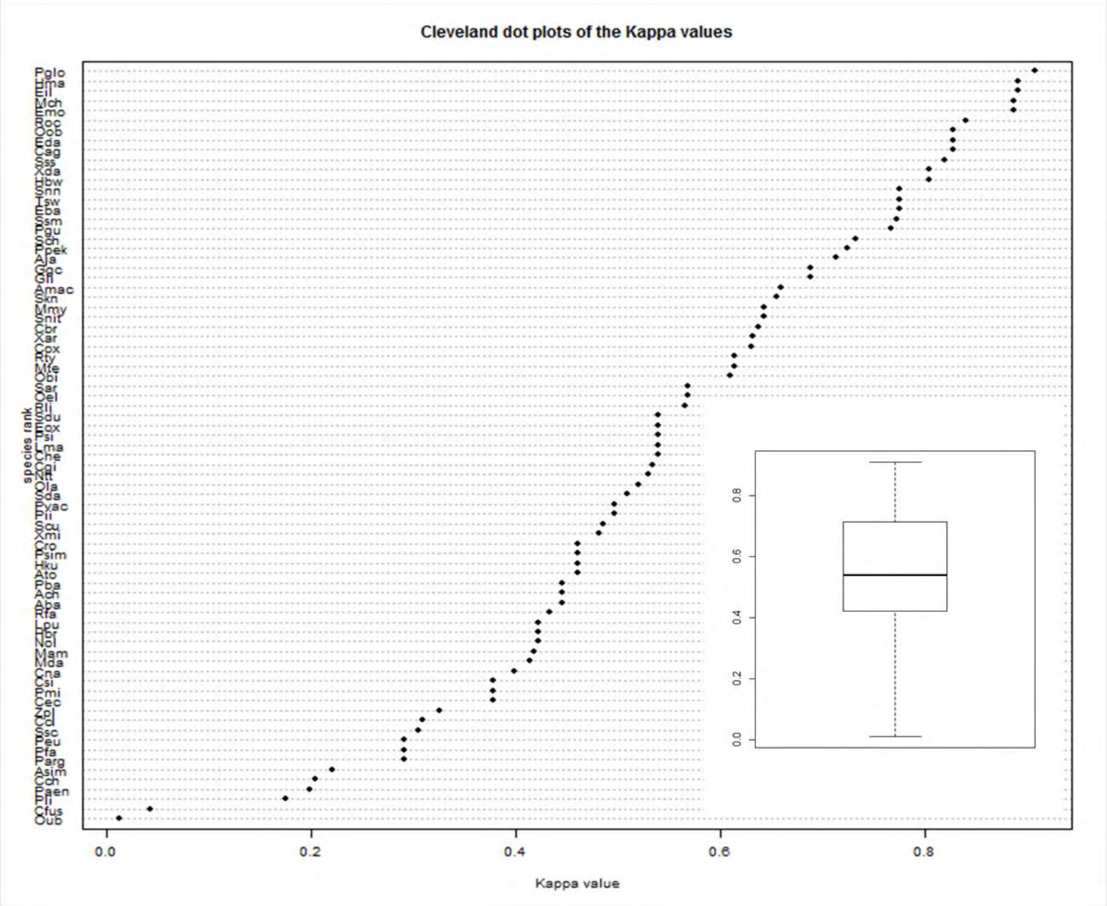


Fig. 16 The species-rank kappa value for the MRT model evaluation. a. Cleveland dot plots of the kappa values **b**. Boxplot of the kappa values

For species richness, a species richness index was used as the response variable. The AUC value (0.5) for the MRT model showed that the model results for species richness is no better than random and thus could be used for fish species richness prediction in Chinese lakes. Simultaneously, precipitation of driest month, maximum temperature of warmest month and lake area were identified to be the most important factors related to fish species richness in Chinese lakes. Lakes with higher precipitation of driest month ($\text{bio14} \geqslant 16.5$ mm) and larger lake surface area (area $\geqslant$ 216.1 km^2) were predicted to support more fish species, approximately 80 species per lake; while the lakes with lower precipitation of driest month (bio14 < 16.5 mm) and lower maximum temperature of warmest month (bio5 < 26 ℃) were predicted to support the fewest fish species, approximately 8 species per lake (Fig. 17).

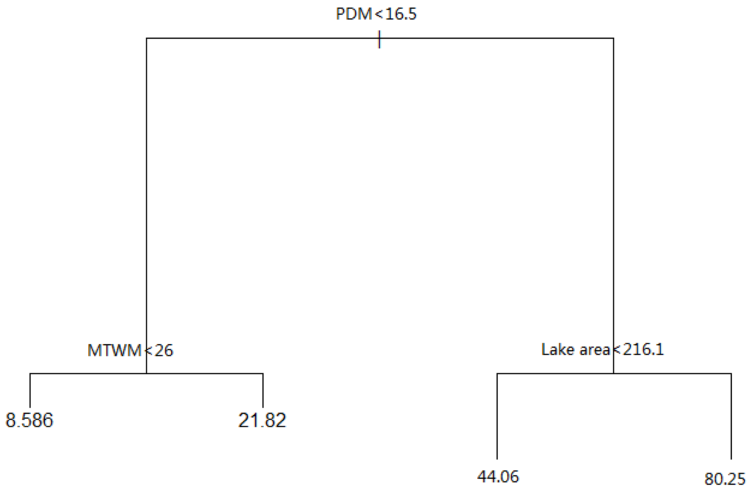


Fig.17 MRT model predicting the fish species richness in Chinese lakes. The AUC value for this model is 0.5

3.3 Ensemble modeling of fish diversity and distributions in China and the uncertainties

Overall, the nine SDMs performed well and showed good capacity on species prediction, as the three different evaluation methods (AUC, Kappa, TSS) and overall accuracy showed significantly high values, with the median values of AUC ranging from 0.680 to 0.891, overall accuracy ranging from 0.85 to 0.90, Kappa ranging from 0.358 to 0.658, TSS values ranging from 0.360 to 0.752 (Fig. 18). Among the entire model techniques, RF always yield the best models since the median AUC value is

0.891, median accuracy score is 0.90, median Kappa value is 0.658 and median TSS value is 0.752, all the values are the highest. Followed by GBM, FDA and GLM respectively, while SRE always gives the relative worst predictive performance with median AUC value of 0.680, accuracy value of 0.85, Kappa value of 0.358 and TSS value of 0.360, which showed the relative lowest values among all the models (Fig. 18). It is worth noting that even if the RF model which performs the best couldn’t give the equal and best predictions for all species, while SRE could also yield very good quality models for some species.

Results from the MANOVA showed that there is no significant difference between the three evaluations used in measuring the performance of nine models ($p > 0.01$).

In order to explore the relationship between species prevalence and model performance, outcomes of all the nine models were taken into account. In the context, only the results of RF model were presented in Fig. 14. Among all the three evaluations, there is a nonlinear relationship between model performance and species prevalence. The species prevalence around 30% could yield more accurately predictive models, while species prevalence below 30%, model performance showed slightly positive relationship with species prevalence (Fig.14).

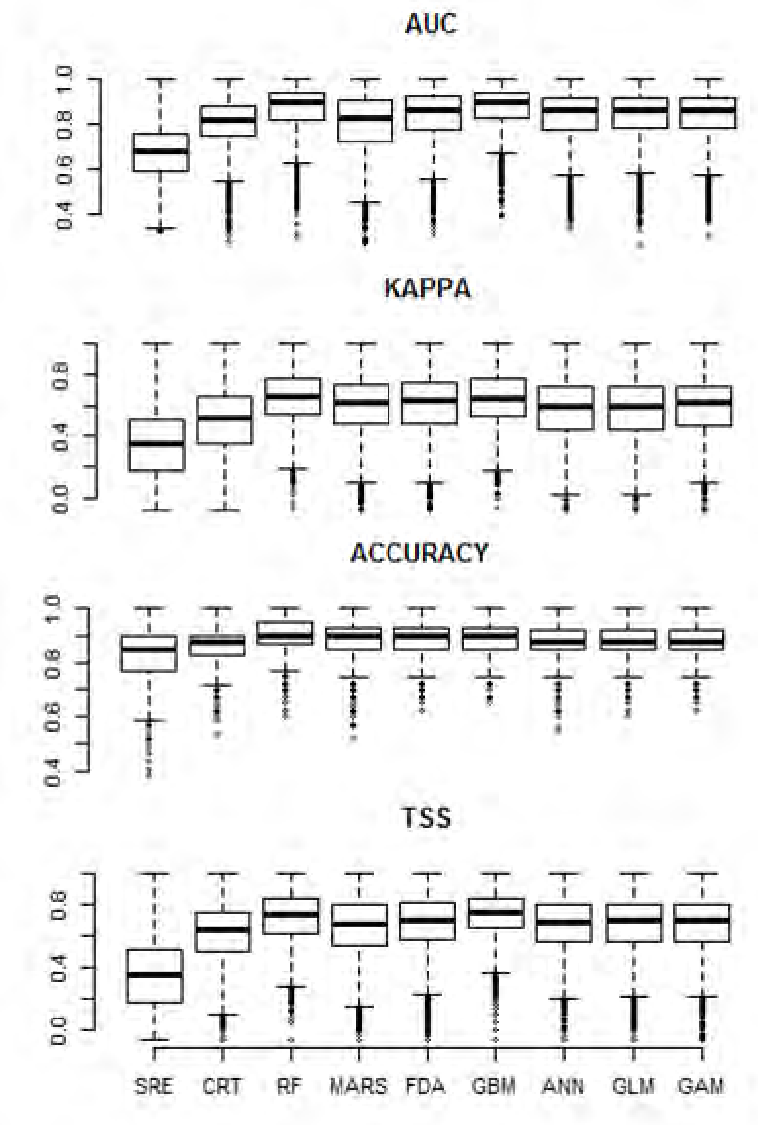


Fig.18 Performance of different SDMs

AUC: Area under the receiver operating characteristic curve; KAPPA: Cohen's Kappa; TSS; True skill statistic; ACCURACY: Overall accuracy. SRE (Surface Range Envelop); CTA (Classification Tree Analysis) ;MARS(Multivariate Adaptive Regression Splines) ; RF (Random Forest) ;FDA (Mixture Discriminant Analysis); GBM (Generalized Boosted Models) ; ANN (Artificial Neural Networks); GLM (Generalized Linear Models); GAM (Generalized Additive Models)

Here in our study, we determined three kinds of ecological range size, altitude, temperature and precipitation ranges. Relationships between species prevalence and species range size were found to be all positively related with all the three environmental range size (Fig.19).

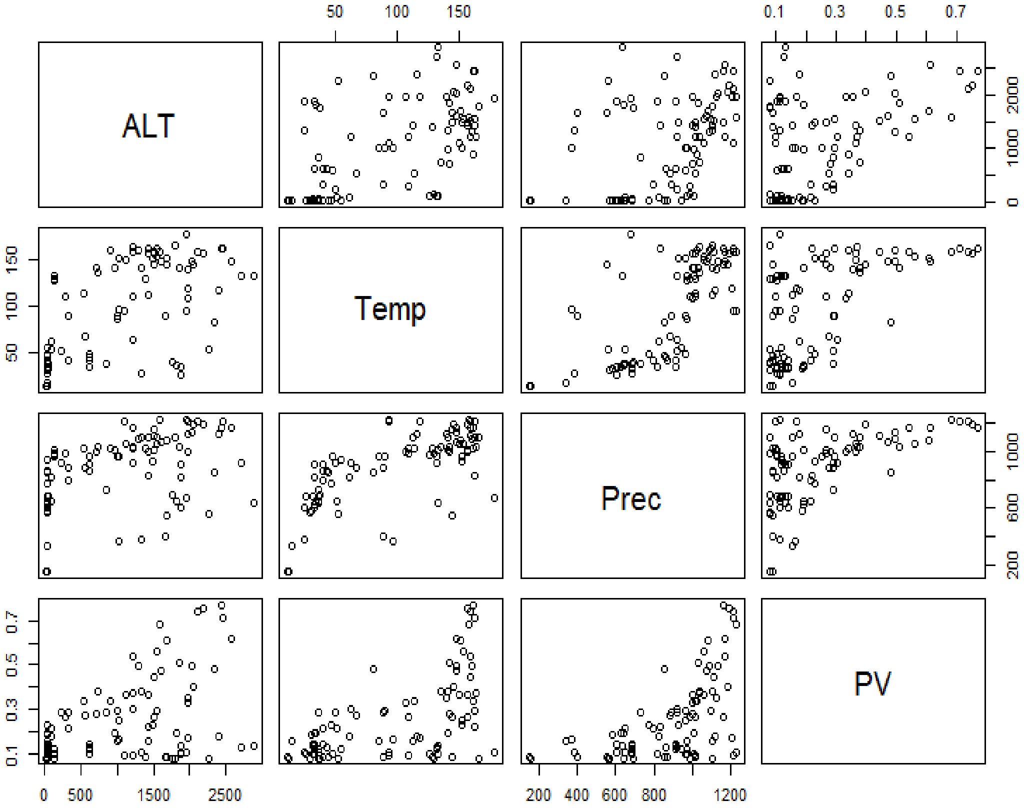


Fig.19 Relationships between species prevalence and three environmental range size

However, when take into account the species ecological range size, the altitude range size and precipitation range size were significantly negative related with the model performance, while unfortunately, temperature range size were found not significantly influence the model performance. Besides, performance of FDA and SRE seems only affected by species prevalence (Tab.3), but not correlated with species characteristics like temperature range size, precipitation range size and altitude range size. Moreover, influence from the uncertainty source on each of the nine models was also showed difference.

Tab. 3 Relations between model techniques and species characteristics

	ALTR		TR	PR		PV	
SRE AUC	NS		NS	NS		0.208	***
CTA AUC	-3.057E-05	***	NS	-1.03E-04	**	0.193	***
MARS AUC	-2.42E-05	**	NS	-9.72E-05	**	0.185	***
RF AUC	-3.43E-05	***	NS	-1.16E-05	***	0.172	***
FDA AUC	NS		NS	-1.28E-04	**	0.233	***
GBM AUC	-3.08E-05	***	NS	1.32E-04	***	0.178	***
ANN AUC	-4.41E-05	***	NS	-1.12E-04	***	0.206	***
GLM AUC	-3.43E-05	***	NS	-6.04E-05	*	0.253	***
GAM AUC	-3.87E-05	***	NS	-7.89E-05	*	0.192	***
SRE KAP	NS		NS	-1.93E-04	*	0.419	***
CTA KAP	-4.43E-05	*	NS	-2.12E-04	*	0.614	***
MARS KAP	-3.73E-05	*	NS	-1.52E-04	*	0.344	**
RF KAP	-3.62E-05	*	NS	-1.73E-04	*	0.283	***
FDA KAP	NS		NS	-1.81E-04	*	0.371	***
GBM KAP	-4.00E-05	*	NS	-2.26E-04	**	0.435	***
ANN KAP	-6.09E-05	***	NS	-1.74E-04	*	0.49	***
GLM KAP	-4.94E-05	**	NS	-1.44E-04	*	0.63	***
GAM KAP	-5.32E-05	**	NS	-1.95E-04	*	0.4399	***
SRE TSS	NS		NS	NS		0.3666	***
CTA TSS	-5.55E-05	***	NS	-2.01E-04	**	0.3626	***
MARS TSS	-4.25E-05	**	NS	-1.78E-04	**	0.268	***
RF TSS	-5.73E-05	***	NS	-2.23E-04	***	0.229	***
FDA TSS	NS		NS	-2.19E-04	***	0.261	***
GBM TSS	-5.20E-05	***	NS	-2.16E-04	***	0.209	***
ANN TSS	-7.89E-05	***	NS	-1.81E-04	***	0.307	***
GLM TSS	-5.66E-05	***	NS	-1.04E-04	*	0.382	***
GAM TSS	-5.81E-05	***	NS	-1.25E-04	*	0.268	***

The model techniques: SRE (Surface Range Envelop); CTA (Classification Tree Analysis) ;MARS(Multivariate Adaptive Regression Splines) ; RF (Random Forest) ;FDA (Mixture Discriminant Analysis); GBM (Generalized Boosted Models) ; ANN (Artificial Neural Networks); GLM (Generalized Linear Models); GAM (Generalized Additive Models) ;**The evaluation techniques**: AUC (Area Under the receiver operating characteristic Curve); KAP(Cohen's Kappa); TSS (True skill statistic);**The species characteristics:** ALTR (Altitude); TR(Temperature Range); PR(Precipitation Range); PV(Prevalence); **NS:** Not Significant

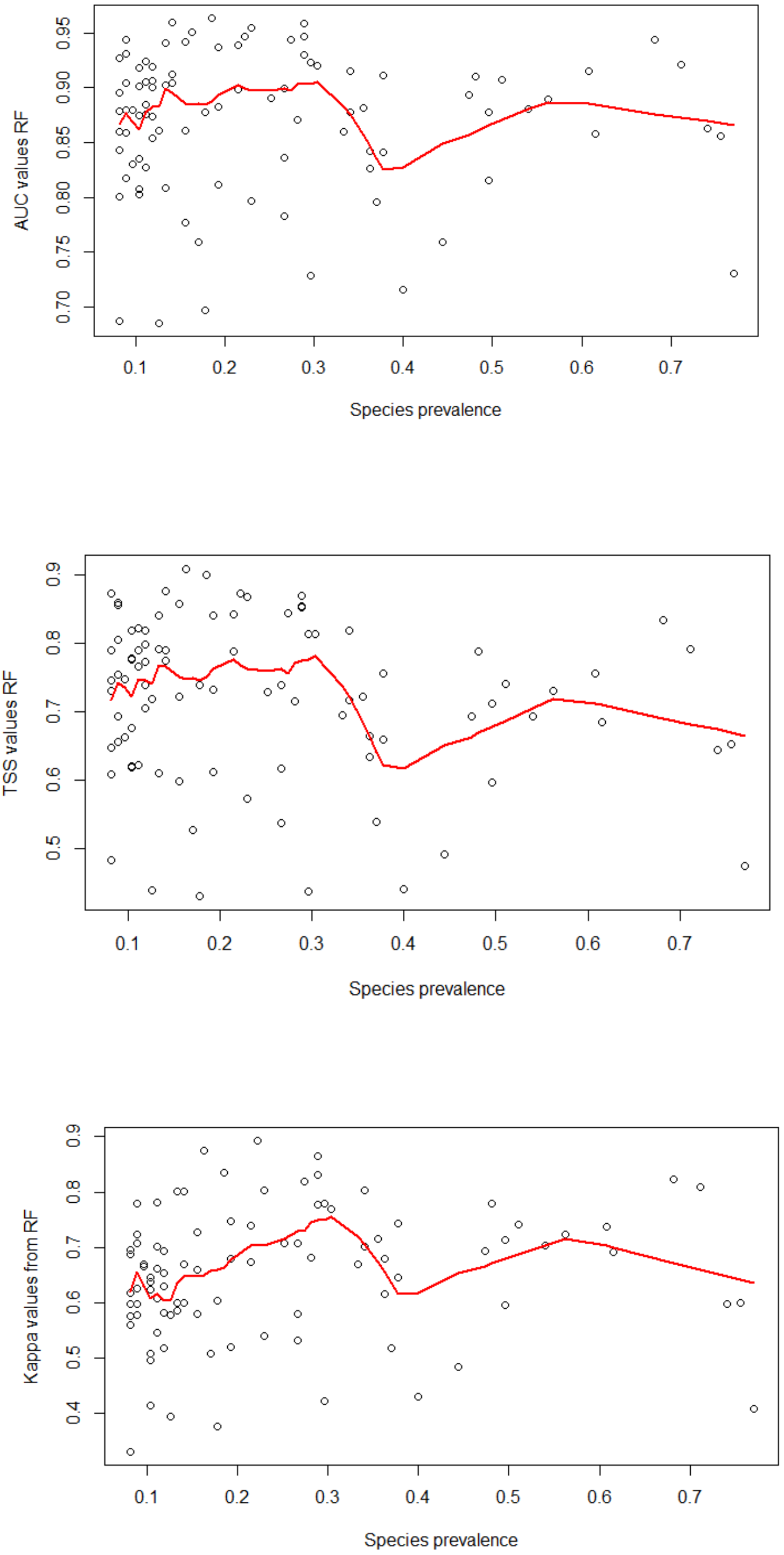


Fig.20 Relationships between species prevalence and model performance

Tab. 4 Basic input and estimated parameters (in bold) for the 23 functional groups of the Bao'an Lake ecosystem model (see details of the abbreviations in 2.2 of the text).

Group Number	Group name	TL	Biomass (t/km^2)	P/B ($year^{-1}$)	Q/B ($year^{-1}$)	EE	P/Q
1	Mandarin fish	**3.138**	1.44	0.78	3.57	**0.554**	**0.219**
2	Snakehead fish	**3.131**	1.38	1.31	3.46	**0.680**	**0.380**
3	Large culters	**3.143**	4.04	0.96	7.88	**0.797**	**0.122**
4	Catfish	**3.080**	0.32	1.42	7.07	**0.704**	**0.201**
5	Common carp	**2.812**	3.26	1.97	9.17	**0.787**	**0.215**
6	Crucian carp	**2.126**	3.80	2.23	17.02	**0.940**	**0.131**
7	Small fishes	**2.151**	18.00	2.30	8.92	**0.785**	**0.258**
8	Silver carp	**2.021**	2.91	1.15	12.24	**0.970**	**0.094**
9	Bighead carp	**2.315**	3.07	1.02	8.62	**0.858**	**0.118**
10	Grass carp	**2.000**	10.42	1.65	12.41	**0.770**	**0.133**
11	Breams	**2.006**	6.17	1.55	18.31	**0.516**	**0.085**
12	Crabs	**2.227**	2.76	2.12	8.48	**0.770**	**0.250**
13	Shrimps	**2.196**	3.97	4.50	24.40	**0.429**	**0.184**
14	Molluscs	**2.000**	99.10	4.30	**86.00**	**0.126**	0.050
15	Oligochaeta	**2.076**	2.01	5.00	**100.00**	**0.959**	0.050
16	Aquatic insecta	**2.056**	0.72	6.20	**124.00**	**0.913**	0.050
17	Microzooplankton	**2.000**	0.70	**32.02**	**640.37**	0.950	0.050
18	Cladocera	**2.016**	0.23	**48.67**	**973.41**	0.900	0.050
19	Copepoda	**2.016**	0.67	**25.29**	**505.81**	0.900	0.050
20	Submerged macrophyte	**1.000**	1627.00	1.25		**0.120**	
21	Attached algae	**1.000**	41.30	250.70		**0.089**	
22	Phytoplankton	**1.000**	2.43	185.00		**0.376**	
23	Detritus	**1.000**	3.45			**0.645**	

3.4 Modelling the lake ecosystem properties and functioning

3.4.1 Basic input and estimates

After balancing the Ecopath model, a series of estimates and analyses were given by the model. The pedigree index (0.50) lies in the upper range (0.16–0.68) from 150 EwE models (Morissette et al., 2006), and the measure of fit is 2.582 for this model, indicating that the input parameters of the model were based on reliable sources and the model was robust with high confidence (Table 4; Christensen et al., 2000).

The basic input with the estimated parameters given by the EwE for the Baoan Lake ecosystem are shown in Table 4, and the diet matrix for each group is presented in Table 5.

For the Bao’an Lake model, the EE values of all groups were less than 1, and most of the P/Q values were between 0.1~0.3, meeting the requirements of a balanced model (Christensen et al., 2005). Generally, the EE values of all the commercial fishery groups were higher than 0.5, and some were even as high as 0.97 (silver carp) and 0.94 (crucial carp). However, the EE of attached algae was extraordinarily low (0.089), followed by submerged macrophytes (0.120) and molluscs (0.126).

3.4.2 Food web structure and trophic analysis

Trophic structure

Fractional trophic levels as suggested by Odum and Heald (1975) were calculated by the Ecopath model. All ecological groups were assigned discrete trophic levels according to Lindeman (1942) with the approach suggested by Ulanowicz (1995). The routine assigns definitional trophic levels (TL) of 1 to producers and detritus and a trophic level of 1 + [the weighted average of preys' TLs] to consumers (Christensen et al., 2005). The fractional trophic levels of all of the 23 groups in Bao'an Lake varied from 1 (primary producer and detritus) to 3.143 (large culters) (Table 4). The primary producers consisted of submerged macrophytes, attached algae and phytoplankton, and the piscivorous fish such as large culter, mandarin fish and snakehead fish occupied the top trophic level in the lake ecosystem.

Two main types of food chains in the Bao’an Lake ecosystem can be discerned from the model: a detrital-based food chain and a grazing food chain. The former chain transferred more energy and matter than the latter one. There were 9055 t km^{-2} y^{-1} matter flowing to trophic level II in the detrital-based food chain, compared with 1333 t km^{-2} y^{-1} in the grazing food chain, although the biomass of the primary producers was extremely high (1671 t km^{-2} y^{-1}) (Fig. 15).

Transfer efficiencies

In order to describe the proportion of energy transferred from one trophic level to the next, all ecological groups in the Bao’an Lake model were assigned to one of eight discrete trophic levels (TL) (Lindeman, 1942). However, only the first five trophic levels were taken into consideration since the flows through TL VI to VIII were extremely small (Fig. 15). The transfer efficiency (TE) of matter is the ratio between the sum of exports and flows predated by the next level and the throughput on the trophic level. For the grazing food chains of the Bao’an Lake ecosystem, the TEs from TL III to TL V were 5.85%, 9.11%, 12.0% respectively. For the detrital food chain, TEs were 1.11%, 13.3% and 10.8% respectively. A mean TE of 8.96% was calculated for the grazing food chain and 8.40% for the detrital food chain. The geometric mean of the trophic transfer efficiency for the Bao’an Lake ecosystem was 8.68%.

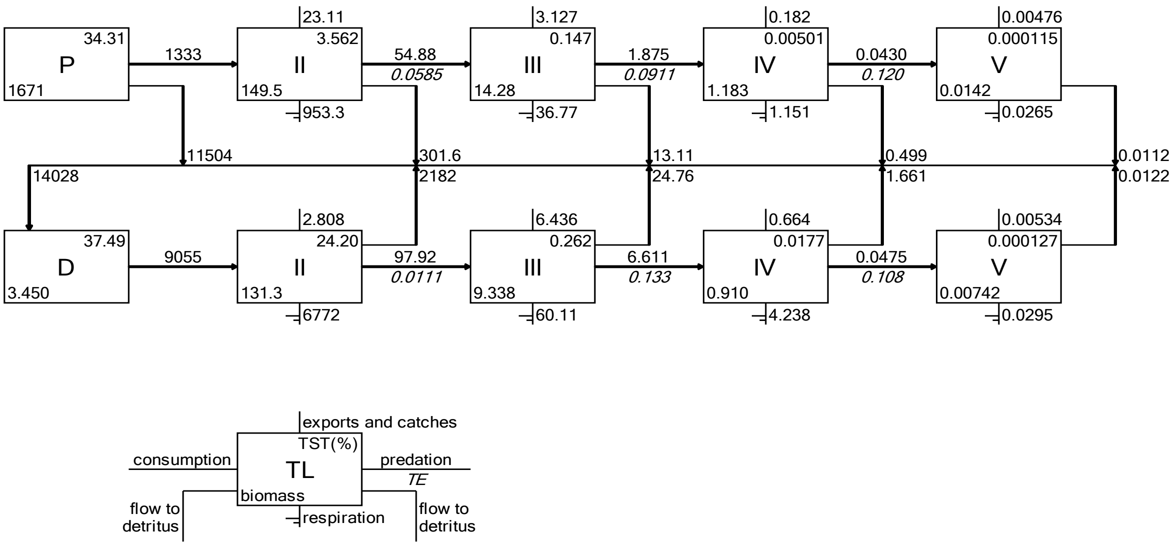


Fig.21 Lindeman spine of Bao’an Lake ecosystem during 1991-1993 (details of the units see Tab. 5).

Mixed trophic impacts

A modified input-output analysis with the procedure “Mixed Trophic Impacts (MTI)” described by Ulanowicz and Puccia (1990) was implemented in the EwE. The MTI describes how any group (including fishing fleets) impacts trophically on all the other groups in an ecosystem. It includes both direct and indirect impacts, *i.e.* both predatory and competitive interactions (Christensen, et al., 2004). From the mixed trophic impact analysis on the Bao’an Lake ecosystem (Fig. 22), piscivorous fish like mandarin fish and snakehead fish had strong negative effects on crucian carp as well as each other. In addition, large culters also exerted strong negative effects on small fish, shrimps, catfish and themselves, while they had moderate negative effects on mandarin fish and snakehead fish.

As the main energy sources in the lake ecosystem, detritus and phyplankton had obvious positive effects on many other functional groups. Fishing had strong negative effects on all the commercial fish and crabs, but was beneficial for the small fish and crucian carp (Fig.22). Nearly all the fish groups had obvious negative effects on forage resources, including forage fish (crucian carps, small fish), zooplankton (microzooplankton, cladocera, copepoda), zoobenthos (molluscs, oligochaeta, aquatic insecta) and algae (attached algae, phytoplankton). On the other hand, the fish groups seemed to have more negative effects on each other mainly due to trophic competition.

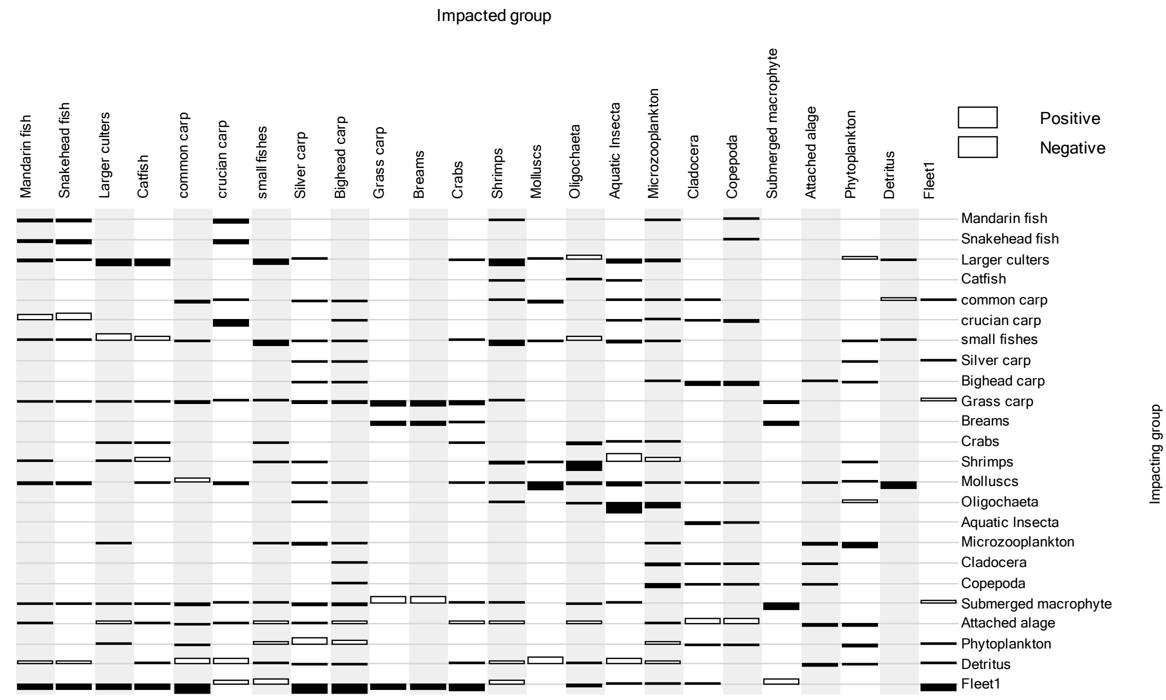


Fig.22 Mixed trophic impacts of Bao'an Lake ecosystem (white blanks above the line represent a positive impact whereas black blanks underneath the line indicate a negative impact, and the heights of the bars are proportionate to the degree of the impacts).

3.4.3 Ecosystem properties and indicators

According to the ecosystem theories implemented by Odum (1969; 1971) and Ulanowicz (1986), a series of indicators that describe ecosystem properties were calculated by the Ecopath to assess the stability and maturity of the ecosystem (Christensen et al., 2005). Summary statistics and flow indices of the Bao'an Lake ecosystem are listed in Table 6.

The total system throughput of the lake ecosystem reached 37418.040 t km^{-2} y^{-1}, of which 28.2% derived from consumption (10549.570 t km^{-2} y^{-1}), 13.4% from exports (5009.366 t km^{-2} y^{-1}), 20.9% from respiration (7827.842 t km^{-2} y^{-1}) with 37.5% (14031.260 t km^{-2} y^{-1}) eventually flowing into detritus. The sum of all production (TP) was 13449.030 t km^{-2} y^{-1}, and the calculated total net primary production (TPP) and the net system production (NSP) were 13449.030 and 5009.368 t/km^{2}/y respectively. Thus, the ratio of total primary production/total respiration (TPP/TR) and total primary production/total biomass (TPP/TB) were 1.640 and 6.993 respectively. The mean trophic level of catch was calculated as 2.262, and the gross efficiency (catch/net primary production) was 0.003 in the lake ecosystem.

Flow indices, including connectance index (CI) and system omnivory index (SOI), were used to describe whether the food web is web-like or linear. The values of CI and SOI in the Bao’an Lake ecosystem were 0.205 and 0.058 respectively. Finn’s cycling index (FCI; Finn, 1976) and Finn’s mean path length (FML) calculated by the model were 9.25% and 2.915 respectively.

Ecosystem information indices, ascendancy (A) and system overhead (O), are derived from information theory as a measure of the average mutual information in a system (Ulanowicz and Norden, 1990). In the Bao’an Lake ecosystem, the values of ascendancy and overhead were 38.7% and 61.3% respectively (Table 6).

Group	Prey \ predator	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	Mandarin fish																			
2	Snakehead fish																			
3	Large culters																			
4	Catfish																			
5	Common carp																			
6	Crucian carp	0.666	0.786																	
7	Small fishes	0.223	0.205	0.790	0.580	0.010							0.150							
8	Silver carp																			
9	Bighead carp																			
10	Grass carp																			
11	Breams																			
12	Crabs																			
13	Shrimps	0.110	0.008	0.193	0.315	0.007														
14	Molluscs		0.001		0.019	0.778		0.132						0.095						
15	Oligochaeta				0.003	0.008	0.007						0.050	0.080						

16	Aquatic insecta			0.001	0.011	0.005	0.008	0.002								0.015				
17	Microzooplankton				0.002				0.006	0.005		0.006		0.005		0.050	0.010		0.016	0.016
18	Cladocera			0.001			0.027	0.008	0.007	0.141				0.003		0.005	0.020			
19	Copepoda						0.081	0.009	0.008	0.164				0.007		0.005	0.025			
20	Submerged Macrophyte										0.977	0.950	0.446							
21	Attached algae						0.050	0.380	0.214	0.200	0.020	0.020	0.149	0.355	0.030	0.380	0.090	0.350	0.537	0.537
22	Phytoplankton						0.022	0.249	0.621	0.383	0.003	0.010	0.005	0.050			0.005	0.150	0.037	0.037
23	Detritus	0.001		0.015	0.070	0.192	0.805	0.220	0.144	0.107		0.014	0.200	0.405	0.970	0.545	0.850	0.500	0.410	0.410
	Sum	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Tab.5 Diet compositions of 23 functional groups in the Bao’an Lake ecosystem mode

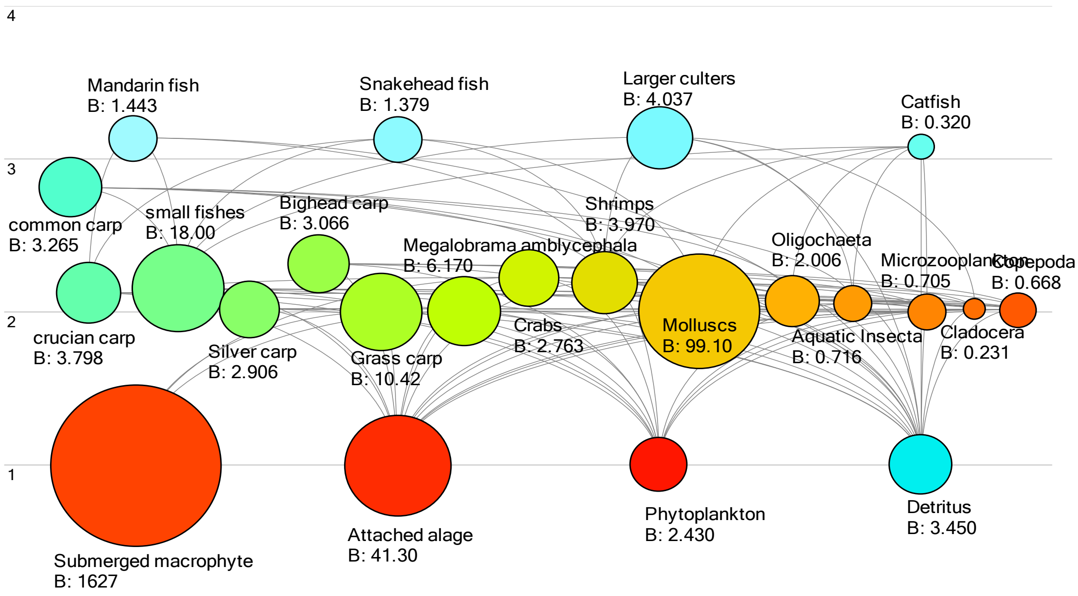


Fig.23 Schematic diagram of trophic flows and food web structure in the Bao’an Lake ecosystem (for biomass the units are t/km^2).

4. Discussion

4.1Fish distribution patterns and the impacts of climate change

The present study is the first to have assessed the patterns of fish species distribution and assemblages as well as their determinants in a large scale in China. Our research provided precisely understanding of fish diversity and distribution patterns, clarified the main drivers, identified the indicator fish species, verified the hypothesis and examined the capacities of MRT model. Rather than some single models that assume species are independent, MRT is a novel tree method that directly determines the assemblages in terms of environment variables. This is suitable and essential for community analysis and produces more creditable results (De’ath, 2002).

Tab.6 Summary statistics of the Bao'an Lake ecosystem properties

Attribute parameters	Value	Units
Sum of all consumption (TC)	10549.570	t/km^2/year
Sum of all exports (TE)	5009.366	t/km^2/year
Sum of all respiratory flows (TR)	7827.842	t/km^2/year
Sum of all flows into detritus (TD)	14031.260	t/km^2/year
Total system throughput (TST)	37418.040	t/km^2/year
Sum of all production (TP)	13449.030	t/km^2/year
Mean trophic level of the catch (TLc)	2.262	-
Gross efficiency (catch/net primary production)	0.003	-
Calculated total net primary production (TPP)	12837.210	t/km^2/year
Total primary production/total respiration (TPP/TR)	1.640	-
Net system production (NSP)	5009.368	t/km^2/year
Total primary production/total biomass (TPP/TB)	6.993	-
Total biomass (excluding detritus) (TB)	1835.691	t/km^2
Connectance index (CI)	0.205	-
System omnivory index (SOI)	0.058	-
Finn's cycling index (FCI)	9.250	% of total throughput
Finn's mean path length (FML)	2.915	-
Ascendancy (A)	0.387	-
System overhead (O)	0.613	-
Ecopath pedigree index	0.500	-
Measure of fit (t*)	2.582	-

4.1.1Fish diversity and assemblages in Chinese lakes

In the present study, five assemblages (**Ia**, **Ib**,**Ⅱa1**,**Ⅱa2**,**Ⅱb**) were ultimately redefined according to the determinants and the fish composition, distinguished from the previous studies: nine administrative regions (Kang et al., 2013) and the five physiographic regions (Zhao et al., 2006; Wang and Dou, 1998). Both the lake richness and fauna diversity showed significant differences in each assemblage: lake richness in the plateaus was clearly lower than the plains, while the diversity of the fauna was the highest.

Assemblage **Ⅰa** mainly consisted of the lakes distributed in the three famous plateaus (Qinghai-Tibet plateau, Yun-Gui plateau, Meng-Xin plateau) in China. Due to the particular natural environments and climatic conditions in plateaus (Wang and Dou, 1998), only one indicator species was found since most are stenochoric species. However, the number of lakes contained in this assemblage was as high as 61. This may partly explain why the lake richness in this cluster was lower (median value = 12sp.) but the fauna diversity was the highest (280 sp.). On the other hand, fish fauna diversity in the assemblage will be even higher with the increasing of research lakes since the species accumulation curve still has not approached the asymptote (Fig. 4). Meanwhile, in these plateaus, the complex of river systems and geographic situation, the specific climate and less effect of human activity made these areas easier for energy to be obtained, and thus could support a high number of species (Wang & Dou 1998; Yang et al., 2004; Zhao et al., 2006; Kang et al., 2013). Zhao et al. (2006) also argued that the low lake richness in plateau was the result of the young geological age for new species colonization and speciation. In the authors' opinion, the results sustained the plausibility of the species-energy hypothesis.

Lakes contained in assemblage **Ⅰb** were mainly distributed in North-east China. In fact, the lakes in this assemblage mainly belong to the Song-Nen plain. However, compared with the low altitude in clusterⅡ, **Ⅰb** was also divided as high altitude area (altitude>=50m). Indicator species found in **Ⅰb** were all especially distributed in northern China water systems with low economic value (Yang et al., 2010). A few previous reports considered that although the weather in north-east China was very cold in winter, the fauna richness in this region is not so low (Zhang, 1999; Zhao et al., 2006). Whereas the fauna diversity (99sp.) was even lower than other assemblages in our study, so was the lake richness (median=22.5sp.). Our conclusion was supported by many field investigations (Ren, 1994; Yang et al., 2010), which indicated that the fish diversity in north-east China was significantly lower than in other areas. Overall, assemblages Ⅰ**b** seems to be the transition region from the plateau to the plain since clusterⅡall contained the lakes in the plains with plain fish fauna.

Generally, lakes contained in cluster Ⅱ could roughly be considered as the eastern plain in terms of lower altitude (altitude<50m), coinciding with the classification of Zhang (1954) and Li (1981). Assemblage Ⅱ**a1** consisted of the largest part of eastern region lakes (38 in total of 48) with *Coilia nasus* (Schlegel) as the sole indicator species. Simultaneously, lakes contained in assemblage Ⅱ**a1** contained moderate lake richness (median value = 55 sp.) and fauna diversity (184 sp.) compared with other assemblages. From the map, we can see that the assemblage occupied two parts, some few lakes distributed in the 3H plain region (Huanghe, Huaihe and Haihe) (Kang et al., 2013), but most of them distributed in the middle-lower Yangtze plain region. The superiority of the natural environment along with the abundance of lake resources has turned the lakes in this area into one of the most important freshwater fishery bases in central China (Cui and Li, 2005).

AssemblageⅡ**a2** only contained 4 lakes that were also distributed in the middle Yangtze River basin. The proper climatic conditions along with ample natural resources in the middle Yangtze River basin showed advantages for supporting more fish species (Liang and Liu, 1995; Cui and Li, 2005; Zhang, 2005), and also concurred with the species-energy hypothesis (Wright, 1983). 59 indicator species were identified in the assemblage. All the indicator species were common and dominant species in the lakes along the middle-lower Yangtze River basin. Assemblage Ⅱb consisted of 6 lakes distributed along the central Yangtze River basin, which contained the highest lake richness (with median value of 88.5 sp.) among all five assemblages. 38 indicator species were thus identified in the assemblage, similar to assemblage Ⅱ**a2**, the indicator species in assemblage Ⅱ**b** were also common plain species in the central-lower Yangtze River. In fact, lakes contained in assemblage Ⅱ**a2** and Ⅱ**b** were either connected to the Yangtze River (*e.g.* Dongting lake, Poyang lake, Junshan lake, Wuhu lake) or isolated from some large lakes (*e.g.* Huanghu lake and Huangdahu lake). Moreover, these lakes used to be connected to the Yangtze River (Shi, 1989) which supported even more common fish species (Fu et al., 2003). Due to the historical river-lake connection, fish composition showed highly homogeneity in these lakes (Ⅱa2 and Ⅱb, no significant difference), and therefore contained most of the indicative species (87.8% of all the indicator species). Therefore, it is no wonder that lakes distributed along the Yangtze River basin have higher lake richness compared with other assemblages. This result was also consistent with recent studies (Zhao et al., 2006; Kang et al., 2013). Consequently, compared with Ⅱ**a2** and Ⅱ**b**, Ⅱ**a1** demonstrated lower lake richness (median value = 55 sp.) mostly because of the disconnection of the lakes and the river (Liu and Wang, 2010; Fang et al., 2006). Furthermore, lake degradation, environmental pollution and irrational fishing modes in these lakes can also decrease the fish richness (Cui and Li, 2005; Fang et al., 2006).

4.1.2 Determinants of fish distribution and assemblages

Overall, altitude, temperature annual range (TAN), precipitation in the driest month (PDM) and the minimum temperature of the coldest month (MTCM) were ultimately identified as the key factors for determining fish distribution and assemblages in Chinese lakes, among which altitude was the most important determinant. Altitude, precipitation and temperature have long been acknowledged to determine the distribution of organisms in various ecosystems, such as plants (Pottier et al., 2013), riverine and marine fish and invertebrates (Bussion et al., 2008; Cheung et al., 2009).

Our results fitted very well with previous research that showed that altitude considerably affects fish distribution in aquatic ecosystems (Amarasinghe and Welcomme, 2002; Yoon et al., 2011; Barradas et al., 2012; Stojkovic et al., 2013).

However, altitude is a complex variable: it can cause direct and indirect effects on fish distribution. Some preview studies have explored the two different influences (Legendre, 1993; Hawkins et al., 2003; Zhao et al., 2006), and have shown that indirect influences were even greater than direct effects, since altitude contained many other factors that could affect fish distribution.

Temperature was an important factor that influenced fish distribution through its affect on fish metabolism (Gillooly et al., 2001), breeding (Mills and Mann, 1985), development and growth (Mann, 1996; Wolter, 2007) and behavior (Taniguchi et al., 1998). Here in our study, two kinds of temperature (TAN and MTCM) have been defined as the important factors which determine the fish distribution, consistent with Crisp (1996) and Mann (1996) who considered these two factors as having great importance in freshwater fish biological requirements. TAN heavily contributed in dividing cluster Ⅰ into two parts, the thermal range appeared to be important in the present study partly because of its vast variability (21.2℃-56.1℃). Since fish are poikilothermal animals they will suffer considerably from the environment temperature change as demonstrated on some European freshwater fish (Buisson *et al.*, 2008; Pont *et al.*, 2005) and American species (Rathert et al., 1999). Although each fish species has a thermal preference and can usually be characterized by a tolerance range of temperatures varying by a few degrees, they are flexible and can adapt to a certain extent by physiological adjustments or behavioral thermoregulation. However, thermal ranges that exceed the proper range will without doubt affect fish lives and distributions (Gislason et al., 2010). This is the explanation as to why the MTCM was a major determinant of the fish distribution and assemblages in our study. In the present study, MTCM acted as the main determinants in cluster Ⅱ. This may be mainly because the fishes distributed in assemblage Ⅱ were adapted to temperate and a subtropical zone climate. Globally, MTCM had also been defined as the most important factor that determines the fish distributions and lives (Rubidge et al., 2011; Yu et al., 2013; Aguilar-Kirigin and Naya, 2013). Other than these two factors, precipitation has always been considered as one of the most important climatic factors in numerous recent studies (Zhao et al., 2006; Buisson et al., 2008; Buisson and

Grenouillet, 2009), because the precipitation could influence the stream flows and hydrological conditions. The hydrology was even more complex and concerns diverse fields such as meteorology, geomorphology, geology or geography which could all affect fish assemblages and distribution. The PDM was also considered as having major impacts on niche shifts of freshwater species (Lauzeral et al., 2011), and thus affect fish distribution. Indeed, the PDM modifies the hydrological conditions of lakes directly, such as water content, water depth and transparency, and nutrients (Wetzel, 2001), which in turn affect fish lives and distributions (Brazner and Beals, 1997; Petry et al., 2003; Mello et al., 2009; Kang et al., 2013). Indeed, besides these factors, there are still many others that could affect and hinder fish distribution and biodiversity, such as land-cover, geographical variables, topological variables, biological invasion, artificial impacts and even net primary production (Park et al., 2006; Buisson et al., 2008; Gevrey et al., 2009; Maloney et al., 2013; Kang et al., 2013). Thus in future studies, more impact factors should be considered comprehensively in order to obtain more powerful results and deeper explanations about the fish distributions and assemblages in lakes.

In conclusion, altitude, precipitation and temperature which could also be attributed to the energy availability explain the fish diversity and distribution patterns across China. However, lake area that have always been considered as the key factor to determine fish dispersal, failed to explain the patterns. All the findings in the current study interpreted that fish species distribution pattern seems to support the species-energy hypothesis rather than the species-area hypothesis.

4.1.3 Potential implications for biodiversity conservation

Our results provide useful perspectives, not only on the current distribution of fish species, but also on the future scenarios under global change. These results revealed important perspectives on species conservation as well as the management of aquatic ecosystems in the future, and thus will be of interest to conservation biologists and environment managers.

Overall, in terms of our findings, the three plateaus that comprised cluster **Ⅰa** should receive high conservation priority on fish conservation since Ⅰ**a** possessed of the highest fauna diversity and richness of endemic fish species. Potential natural reserves should be considered for the conservation of endemic and endangered fish species. Indeed, the construction of natural reserves has long been considered as one of the effective approaches to conserve fish resources (Park et al., 2003; 2004; Zhao et al., 2006; Kang et al., 2013). These conservation strategies are mostly focused on the endemic fish since they are abundant in Chinese lakes and rivers (Park *et al.,* 2003; Fu *et al.,* 2004; He *et al.,* 2011). However, we still cannot ignore the impacts of invasive species since they are considered as one of the greatest threats to global biodiversity (Butchart et al., 2010; Vorosmarty et al., 2010). This is especially true for freshwater ecosystems that are particularly vulnerable to biological invasions (Cucherousset and Olden, 2011).

Otherwise, in the authors' opinion, the most important plain region, Ⅱ**a1,** that contains most of the plain fish fauna together with the highest lake richness needs extra consideration, since Ⅱ**a1** contains most of the lakes which provide major commercial freshwater fisheries in central China (Cui and Li, 2005). As conservation should be carried out in a networked region rather than a single reserve (Bonn and Gaston, 2005), different conservation strategies should be proposed aimed at different objectives. By contrast, with Ⅰ**a**, the conservation strategies here should be to upgrade the fisheries to an eco-friendly modes, to improve the fish diversity and the health of the ecosystem (Guo et al., 2013). Moreover, one suggestion for the conservation of fish diversity in this area was the connection of the lakes with the Yangtze River (Fu et al., 2003; Liu et al., 2010).

4.2 Predicting fish species richness and distributions

Numerous studies have evaluated the predictive performance of different modeling techniques over the last decade. In the current study, the capacity of MRT models in predicting fish species assemblage patterns and species richness patterns based upon geographical and climatic variables was examined in Chinese lakes.

Generally, we emphasize that MRT was a robust modeling technique for species prediction under global changes; ecologically, our results highlight the significance of climatic and geographical variables in determining fish assemblages and distributions in aquatic ecosystems.

4.2.1 Model performance and technique assessment

MRT is used as the main approach for species distribution prediction. A 100 cross-validation test was employed here to strike a balance between explanatory and predictive power and to obtain a more “honest” assessment of the model, with avoid the over-fits of the multivariate regression tree analysis. Thus the minimum cross-validated error tree with six nodes was selected in this study (De’ath and Fabricus, 2000).

AUC, overall accuracy and Cohen’s Kappa were employed to better evaluate the model performance and assess the possibilities of MRT modeling for species distribution prediction. AUC is always considered as one of the most useful and popular performance measures in common ecological model studies (e.g. Manel et al.,1999; 2001; Guisan and Zimmermann, 2000; Pearce and Ferrier, 2000; Luck, 2002; Gibson et al., 2004; Araujo et al., 2005; Luoto et al., 2006; Buisson et al., 2008a; 2008b; Buisson et al., 2009; Grenouillet et al., 2011). It is considered to be the strictest performance measure by some researchers, because it is an independent threshold measure of quality (Fielding and Bell, 1997; Buisson et al., 2008a). AUC scores of 0.5 indicate that a model has no discriminatory ability, while a score of 1 indicates that presence and absence are perfectly discriminated. In our study, AUC values of the indicator species varied from 0.53 to 0.99 with an average value of 0.87, indicating that almost all the species seemed to be perfectly predicted by the MRT model. The same was observed for overall accuracy, where the results varied from 0.61 to 0.98 with an average of 0.84, also indicating that all the species could be accurately predicted by the MRT model with strong discrimination.

Some researchers criticize that AUC and overall accuracy may overestimate model performance (Lobo et al., 2008; Peterson et al., 2008; Jimenez-Valverde, 2012). However Cohen’s Kappa is considered as one of the good solutions to the overestimate problems observed with AUC and overall accuracy (Liu et al., 2011) as the Kappa index can provide a proportional accuracy for predicted presence and absence (Cohen 1960), and provide a robust evaluation of a model’s performance. Kappa has already been widely used in recent ecological studies (Manel et al., 2001; Liu et al., 2005; Virkkala et al., 2005; Buisson et al., 2008a; Gevrey et al., 2009). Cohen’s Kappa values in the current study varied from 0.012 to 0.91 with a mean value of 0.55. A mean Kappa value of 0.55 also indicates that most of the species were accurately predicted by the MRT model, except the three species (*Opsariichthys uncirostris bidens*, *Clarias fuscus* and *Paracheilognathus iimberbis*) which have lower kappa values between 0 and 0.2. There are many factors, such as lower prevalence, spatial autocorrelation, species attributes, environmental range size and, potentially, the model technique, which may result in the observed low accuracy of prediction for these three species.. In the current study, the results may mainly be affected by low occurrence and the environmental range of the three species. Normally for some rare species, poor performance measures would be obtained. Buisson et al. (2008a) found that rainbow trout was the rare species which obtained the low Kappa value. Actually, this phenomenon is common in some previews studies (Manel et al., 2001; Liu et al., 2005; Gevrey et al., 2009). The species’ environmental range could also be one of the main factors which drive the uncertainty of the prediction of the three species in this study, as numerous studies validate that species with a smaller range can be better predicted than species with a larger environmental range (Hernandez et al., 2006; Grenouillet et al., 2011).

Overall, MRT models have been evaluated to be effective and robust enough for species distribution prediction from the ensemble evaluated measurements, even though the kappa values were marginally affected by some rare species. In conclusion, the model performance values provided optimistic estimates of the true predictive capability of MRT modeling (Araujo et al., 2005). Therefore, MRT is a robust and appropriate approach for modeling similar datasets, as it can handle complex ecological data with mixed change and high-order interactions (De’ath, 2002; Davidson et al., 2010).

Currently, although numerous studies focus on defining fish distribution patterns in conjunction with the correlated environmental factors, most are based on the species-specific models (LM, GLM, GAM, CART, RF) which typically assume that each species is independent from the others that occur at the same location (Larsen and Speckman, 2004; Guisan and Thuiller, 2005). However, in natural ecosystems, one species always coexists with many other species to form a community, within which all species are sensitive to resource competition (Wetzel, 2001). Thus, species-specific models fitted to each species would ignore the inter-specific relationships that would provide alternative and variable information about species distribution (Yin, 1990), while the prediction approach at the community scale could overcome the disadvantages of the traditional species-specific focus (Joy and De'ath, 2004; Olden et al., 2006).

4.2.2 Prediction and determinants of fish species assemblages in Chinese lakes

From the validated MRT model, we conclude that ALT (altitude), PDM (precipitation of driest month), TAR (temperature annual range) and AMT (annual mean temperature) were ultimately identified as the key factors in predicting the fish species distribution in Chinese lakes, among which altitude was the most important determinant. In fact, altitude, precipitation and temperature have long been acknowledged as determinants of the distribution of organisms, such as plants (Prasad et al., 2006; Pottier et al., 2013) and riverine and marine fish and invertebrates (Bussion et al., 2008a; Cheung et al., 2009).

Altitude generally impacts distribution and life history traits of organisms globally (Amarasinghe and Welcomme, 2002; Zhao et al., 2006; Yoon et al., 2011; Barradas et al., 2012; Stojkovic et al., 2013). As a complex variable, it can have direct and indirect impacts on fish distribution. Previous studies that have explored these impacts (Legendre, 1993; Hawkins et al., 2003; Zhao et al., 2006), indicated that indirect impacts were even more significant than direct impacts, since so many other factors that could affect fish distribution also vary with altitude.

Without doubt, temperature is an important factor that influences fish distribution through fish metabolism (Gillooly et al., 2001), breeding (Mills and Mann, 1985), development and growth (Mann, 1991; Wolter, 2007) and behavior (Taniguchi et al., 1998). In our study, two kinds of temperature statistics (TAR and AMT) were defined as important in the determination of fish distribution, which is consistent with Crisp (1996) and Mann (1996) who considered these two factors as of high importance for the biological requirements of freshwater fish. The influence from TAR in the present study may be partly due to the vast variability of the temperature (21.2 - 56.1℃), as fish are poikilothermal animals and therefore highly sensitive to changes in temperature. Some European freshwater fishes (Buisson et al., 2008a; Pont et al., 2005; Crisp, 1996; Mann, 1996) and American species (Rathert et al., 1999) were also demonstrated to be great impacted by the TAR. Although each fish species has a thermal preference characterized by a tolerance range of temperatures, they are flexible and can adapt to a certain extent with physiological adjustments (Johnson & Kelsch, 1998) or behavioral thermoregulation (Heggenes et al., 1993). However, thermal ranges exceeding the normal range will impact fish populations and distributions (Gale et al., 2013; Gislason et al., 2010). Extremely low temperatures may affect metabolism, breeding, growth, behavior, and thereby fish distributions (Ruibidge et al., 2011; Yu et al., 2013; Aguilar-Kirigin and Naya, 2013).

Apart from altitude and temperature, precipitation has been considered as one of the most important climatic factors in numerous recent studies (Zhao et al., 2006; Buisson et al., 2008a; b; Buisson and Grenouillet, 2009). Precipitation may impact stream flows and hydrological conditions, while hydrology is even more complex and concerns diverse factors such as meteorology, geomorphology, geology or geography, which could all affect fish assemblages and distribution. The PDM is also considered as having a large impact on niche shifts of freshwater species (Lauzeral et al., 2011), and thereby affecting fish distribution. Indeed, the PDM can modify the hydrological conditions of lakes directly, such as water content, water depth and transparency, nutrient supplements (Wetzel, 2001), which in turn affect fish populations and distributions (Brazner and Beals, 1997; Petry et al., 2003; Mello et al., 2009; Cheng et al., 2012; Kang et al., 2013).

The set of the eight variables used in this study was relatively successful in predicting and explaining the fish assemblages using the MRT model. However, a number of factors have been identified as potentially affecting fish distribution and diversity globally, such as land-cover, geographic variables, topological variables, biological invasion, artificial impacts and net primary production (Park et al., 2006; Buisson et al., 2008a,b; Gevrey et al., 2009; Maloney et al., 2013; Kang et al., 2013). Further research therefore requires the inclusion of such factors in order to obtain more powerful explanations. More important, as we live in a changing world, predictions of species distributions based on possible future scenarios are more desirable than the explanation of the current situation. Thus, more efforts should be made to improve the understanding of the potential distribution and impacts of fish species associated with the global climate change.

4.2.3 Prediction and determinants of fish species richness in Chinese lakes

Species richness is one of the most important biodiversity indicators in ecology. At this level, our research revealed that fish species richness was highly related to precipitation of driest month (PDM), maximum temperature of warmest month (MTWM) and lake area (A). As seen previously, fish species distributions are highly related to precipitation and temperature, it therefore follows that species richness should also be affected by precipitation and temperature. Results also showed that higher fish species richness is predicted in lakes with sufficient precipitation and a larger surface area. This is in agreement with the actual situation in China where lakes distributed along the middle reach of the Yangtze River, such as Dongting lake, Liangzi lake, Hongze lake, Poyang lake, have a higher fish richness as they are exposed to an appropriate temperature, abundant rainfall and are lakes with relatively larger areas (Cui and Li, 2005). However, lakes distributed in parts of Qinghai-Tibet, Meng-Xing and north-east China have relatively lower fish richness due to the lower precipitation and lower temperature (Zhao et al., 2006). In fact, several hypotheses have been presented on species richness patterns, among which the species-area hypothesis and the species-energy hypothesis (Wright, 1983) are the most popular. In this study, we found that PDM, MTWM and lake area best explained the variability of fish species richness patterns in China. These results therefore support the plausibility of both hypotheses for the fish of Chinese lakes. Our results support the species-energy hypothesis, which claims that environments with higher energy availability could support more species, as well as supporting the hypothesis that large surface areas could support more fish species, although in this study lake area was not the highest contributer to the prediction of species richness. The lake area effects may be overshadowed by the large variations of the climatic variables in our study. Our findings were generally consistent with some studies in terrestrial ecosystems (Wright et al., 1999; Hawkins et al., 2003a), and freshwater ecosystems (Zhao et al., 2006). Some other studies, with a smaller scale focus, also argued that fish species richness was highly related with water depth (Cheng et al., 2012), habitat (Petry et al., 2003) and other organisms (Xie et al., 2001) in lakes.

The results of the current study may provide a basis for future research using MRT as well as contributing to conservation of fish biodiversity under global changes.

In conclusion, our research focuses on the understanding of fish species richness and distributions patterns as well as the potential driving factors behind these parameters in Chinese lakes. The recursive partitioning and regression tree models MRT were used and thus be examined. Simultaneously, we have drawn regarding the relative roles of the environmental and climatic variables in driving the lake fish species distribution and richness patterns.

4.3 Ensemble modeling of fish species distribution and the uncertainties

It is important to understand why and how species distribution models perform differently for different species before using the model predictions to make conservation decisions. Up until now, our research is the first study which quantified the uncertainties and impacts comprehensively from model techniques, environmental range size and species prevalence on the SDM performance in lake ecosystems in so large geographical scale. Overall, our results have compared the performance of nine widely used species distribution models; clarified that evaluation measures won’t influence the model outcome; confirmed that species prevalence and environmental range size can strongly affect model performance and ecological predictions; and verified the hypothesis that specialized species could be better predicted than generalized species is plausible in aquatic ecosystems.

Among all the statistical techniques, RF was found to be the most reliable model for species prediction, while SRE predicted the worst. However, the predictions from different models varied a lot, even if for one given species, outcomes of prediction may vary from model to model. In view of that each predicts models relied on different mathematical functions, SDM will give the variety of results without doubt. Up until now, numerous of studies have compared the accuracy and performance of the predictions from different statistical techniques (Elith et al., 2007), and confirmed that results derived from different model techniques or different model-building assumptions can occasionally differ grossly (Thuiller, 2003; Luoto et al., 2004). Nevertheless, it was not so surprise since RF model gives the predictions by generating thousands of trees and aggregated with an average (Breiman, 2001), and the algorithm allow the model to avoid over-fit, this procedure could improve the predictive performance and reduce the variance (Elith et al., 2008). Thus, RF could be a robust technical modeling for species distribution prediction (He et al., 2010; Cheng et al., 2012; Grenouillet et al., 2011). Actually, plenty of publications have noted the algorithm which Random Forest relied on, they thus present the ensemble modelling framework which aggregated several single models and given the average or consensus results (Araujo and New, 2007). Several former studies have verified that among plenty of mathematical models, only RF could show the equal performance with the average outcomes of several model techniques. Therefore ensemble modelling was also regarded as the best solutions to reduce the single model uncertainties and bias (Grenouillet et al., 2011; Buisson et al., 2010b).

Considered that some authors argued that AUC statistic may be biased for the species that occupy a small proportion of the study area (Lobo et al., 2008), in current research, three different measurements have been adopted in order to reduce the potential bias may emerge from single measurement. However, in current study, among all the evaluation results from AUC, TSS and Kappa, there is no significant difference between the three evaluate methods, which interpret that the use of evaluation measures didn’t interrupt the outcomes of statistical models, we can thus verified that the relationship between species characteristics and model performance was not artificial associated with use of accuracy measures (Newbold et al., 2009).

In some previous studies, publicized effects of species prevalence on model performance showed complicated with both positive and negative relationships (Luoto et al., 2005; Brotons et al., 2004; Mantel et al., 2001). In current study, we found that species with prevalence around 30% could yield more accurate model performance, however, the relationships between species prevalence and model performance was nonlinear, model performance increased slightly with species prevalence values below 0.3. Actually, these results were completely in consistent with the former research which argued that species with high prevalence will occupy large areas with variety of habitat environments which would produce more overall errors (Segurado and Araujo, 2004; Luoto et al., 2005). Actually, species prevalence has long been reported which will affect model performance, however few ecologist take into account the species prevalence when they make the species distribution modeling (Fielding and Bell, 1997). Manel et al. (2001) revealed that model performance were highly associated with species prevalence and thus recommend that species distribution model with presence-absence data should take more attention of species prevalence.

Overall in our study, the performance and ecological predictions of the species distribution modelling was negatively associated with geographical and environmental range size, such as altitude range size and precipitation range size. Actually, A general pattern has been widely acknowledged is that species with smaller geographical extent and strict ecological requirements (i.e. specialists) yield models with higher accuracy than those with larger areas of occupancy (i.e. generalists), this hypothesis have been verified in various of ecosystems, among the species as butterfly, insect, reptile, bird, and mammal (Stockwell and Peterson, 2002; Brotons et al, 2004; Segurado and Araujo 2004; Kadmon et al., 2003; Seggurado and Araujo, 2004; Hernandez et al., 2006; Tsoar et al., 2007; Franklin et al., 2009). Our research has verified this hypothesis to be plausible to fish species in lake ecosystems. To our knowledge, the current research is the first research which takes into account fish species in lakes in so large scale, therefore, it would not only contribute to understand the uncertainty of species distribution models, but also help to improve the quality of fish species prediction which will ultimate benefit to the fish biodiversity conservation and biodiversity management all over the world.

When considering about the reason why specialized species could yield more accurate model prediction than generalized one, some authors argued that species with larger distribution area contained discrete populations that show different response to the environment and thus decrease the accuracy of the model prediction (Stockwell and Peterson, 2002; Brotons et al., 2004). Some others explained these as the difference of species' niche width, since former studies have found that species with narrow, well-defined niches which also have better-defined climate and habitat requirements could be better modeled than those with broader niches (Boone and Krohn, 1999; Pearce et al., 2001; Kadmon et al., 2003). But our data couldn't well support and explain the species niche hypothesis, further study should take more factors into account in order to better understand the mechanism of uncertainty in species distribution models from species attributes and characteristics. Last but not least, McPherson and Jetz (2007) attributed the effects of species' ecological characteristic on SDMs were influenced by the available data quality or by making it difficult to statistically capture the relationship between the species distribution and environmental conditions. Following these hypothesis, we can conclude that endemic species were modeled more accurately than non-endemic species. Given all of these, greater consideration should therefore be taken on the predictions of fish species with larger elevation range size and precipitation range size in China under the future impacts of global climate change, since the projections may somehow unreliable for the conservation and management purpose. On contrary, ecological predictions for the endemic fish species which showed better defined climate and habitat requirements should be more accurately. These results will benefit a lot to the conservation of biodiversity for fish species in China, since there are numerous of endemic and specialized fish species lived in the lakes across China, and urgent stages were planned to conserve the fish species based on the prediction results (He et al., 2010). However, it is worth noting that in our study, temperature range size didn't show a significantly affect on the model performance, this founding contradicted some former results in fish species (Grenouillet et al., 2011). In the author's opinion, we contributed this to the large range size of the geographical factor, and the effects of the temperature range may be hidden since that altitude is a complicated factor.

To conclude, we strongly recommend paying more attention on the following aspects when predicting the potential impacts of global climate change on fish biodiversity and distribution using species distribution models. First of all, more suitable model should be implemented and more statistical model techniques should be taken into account in the ensemble model framework; Second, high quality fish presence-absence dataset should be compiled; Last but not least, species ecological and geographical characteristic should be taken into account when predicting the species distribution and diversity. Following these recommendations, predictions and projections of fish species distribution based on the results of SDM could be more reliable and therefore robust enough for management and conservation decisions made under the future impacts of global climate change.

4.4 Effects of anthoropogenic activities on the lake ecosystem properties and functions

The current study contributes to the establishment of a mass-balance model to describe the features of food web structure and ecosystem properties in a shallow macrophtyic lake with the objectives of claiming the effects of artificial activities as fish stocking on the lake ecosystem functions and properties. The results could also be viewed as guiding the development of an eco-friendly fishery and the protection and restoration of submerged plants, since the Ecopath with Ecosim model is a systematic and comprehensive model. This model seems to be the first one established in one of the thousands of shallow macrophtyic lakes distributed along the middle reaches of Yangtze River basin. It provides in-depth knowledge of shallow macrophtyic lakes and thus can help to direct us more sustainably in the development and utilizition of lake resources.

In the Bao'an lake ecosystem, the EE values for most fish groups were high, especially for some commercial and stocked fish groups, such as silver carp, bighead carp and grass carp, bream, common carp, mandarin fish and large culters. Additionally, some prey fish groups, such as small fish and crucian carp also showed extremely high EE values. This suggests that commercial fish species are suffering from overexploitation while prey fish suffer from a combination of pressures from predation by piscivores as well as humans, which has lead to a sharp decline in fishery resources (Hu and Huang, 1991). In former times, traditional Chinese carp like silver carp, bighead carp and grass carp were considered to be the main economic stocking species which explains why these fish groups have reached such high biomass in the lake. Another notable feature in the macrophtyic lake is that the biomass of small fish (18 t km^{-2}) and molluscs (99.10 t km^{-2}) is much higher than in other lakes, possibly because submerged plants provide sanctuary for small fish and an attachment matrix for molluscs (Li et al., 2010; Xie et al., 2006), and thus these resources have not been utilized sufficiently due to the traditional fishery model.

In contrast, the EE values of primary producers, such as submerged plants, attached algae and phytoplankton were as low as 0.12, 0.089 and 0.376 respectively, while the gross efficiency (fishery catch/net primary production) was only 0.003, even lower than Taihu lake (0.0087; Li et al., 2009). This means that primary production was not efficiently utilized and thus very little was transferred into fishery products. Nevertheless, the eficiency of Bao'an Lake is still higher than the global average (0.02%) (Christensen et al., 2005). Large parts of primary production (89.62% of total primary production, about 11504 t km^{-2} y^{-1}) flowed into detritus. Additionally, the food resource of the Bao'an Lake Ecosystem was derived mainly from detritus

(87.17%) with little from primary production (12.83%), indicating that the Bao’an lake ecosystem depended much more on a detritus-based food chain than a grazing food chain, which is also consistent with the suggestion by Odum (1969) that a mature system may depend more on the detrital pathway. The fact that the Bao’an lake food web relies on two energy sources may enhance the plasticity and resilience of the ecosystem (Fetahi et al., 2011). According to the MTI analysis, the two main food sources in the Bao’an lake ecosystem, i.e. detritus and primary producers (especially attached algae and phytoplankton), had significant positive effects on other groups. Other important information from the MTI is that fishing pressure may have more negative impacts on commercial fish groups but will be better for small fish resources, consistent with the fact that fish tend to miniaturization (Cao et al., 1991). The MTI also showed thatfishing pressure exerted stronger impacts on most functional groups than predation or competition (Christensen et al., 2004).

The mean transfer efficiency among different trophic levels of the ecosystem was 8.68% in our study, slightly lower than the 10% assumed by Lindeman (1942), but lying in the accepted range of EE values reported in the published literature (Libralato et al., 2008, Pauly and Christensen, 1995). The low transfer efficiency mainly results from transfer from TL II to TL III in the food chain. From the model, it is evident that a large amount of mollusc biomass was not utilized efficiently.

As previously mentioned, no such model has been established for the lakes along the middle reaches of the Yangtze River, although there are some preliminary studies in some large lakes nearby: Taihu lake (Li et al., 2009) and Gehu lake (Jia et al., 2012) (Table 7). A comparison with these three lake ecosystems on some key parameters from network analysis would be helpful for accurately positioning and characterizing the Bao’an lake ecosystem (Table 6). Notably, the total system throughput in Bao’an Lake (37418.04 t km^{-2} y^{-1}) is much higher than in Taihu Lake (13586 t km^{-2} y^{-1}) or Gehu lake (12131.76 t km^{-2} y^{-1}). This is mainly because the biomass of primary producers, such as submerged plants accompanied by attached algae was much higher than in the other two lakes, indicating that Bao’an lake is typically a lake dominated by submerged plants.

According to Odum (1971), the ratio of total primary production to total respiration (TPP/TR) is an important measure of ecosystem maturity; ecosystems with the values much higher or lower than 1 are thought to be immature, while only those with TPP/TR ratios approaching 1 are considered to be mature. In our study, the value of the TPP/TR ratio was 1.640, much lower than the 3.85 in Taihu or 2.761 in Gehu (Li et al., 2009; Jia et al., 2012), suggesting that the Bao’an lake ecosystem was more mature compared with the other two lakes. Additionally, as modified from Odum (1969), the lower TPP/TB ratio also characterizes a mature ecosystem. The ratio in this study was 6.99, between the immature ecosystem (Taihu Lake, 11.66) and the mature ecosystem (Gehu lake, 1.76). In Bao’an Lake, ascendancy (38.7%), a measure of ecosystem growth and development, is much higher than in Taihu Lake (25.9%) or Gehu lake (33.2%) which also suggests a mature system.

Tab.7 Comparison of ecosystem attributes in different shallow lakes in China

Parameters	Taihu Lake[a] (1991-1995)	Gehu Lake[b] (1986-1989)	Bao’an Lake[c] (1991-1993)
Mean trophic level of catch (TLc)	2.92	2.78	2.26
Total system thoughput (TST)	13586	12131.76	37418.04
Total primary production/total respiration (TPP/TR)	3.85	2.761	1.64
Total primary production/Total biomass (TPP/TB)	11.66	1.76	6.99
Ascendancy (A)	0.259	0.332	0.387
Connectance index (CI)	0.206	0.208	0.205
System omnivory index (SOI)	0.042	0.086	0.058
Finn’s cycling index (FCI)	11.58	14.76	9.25

a. Li et al. (2009)

b. Jia et al. (2012)

c. Present study

FCI represents the fraction of an ecosystem’s throughput that is recycled compared to total system throughput (Finn, 1976; Table 3). The value of 9.25% for Bao’an Lake is much lower than in the other two lakes (Taihu lake, 11.58%; Gehu lake, 14.76%). Meanwhile, CI and SOI are two other important indices which describe system maturity since the food chain is expected to change from linear to web-like as the system matures (Odum, 1971). However, in Bao’an Lake, we can see that the value of CI and SOI (0.205 and 0.058 respectively) were both relatively low. Although these values are nearly in the same order of magnitude as some mature systems such as Gehu (Jia et al., 2012), Qiandaohu (Liu et al.2008) and Taihu lakes(Li et al., 2009), there is still a large gap (especially SOI) with some other mature systems such as Hayq lake (SOI=0.224; Fetahi et al., 2011), Lake Toya (SOI=0.12; Hossain et al.,2010) and Lake Kuvi (SOI=0.148; Villanueva et al., 2008).

Consequently, the high values of TPP/TR and ascendancy and the low values of TPP/TB, and the more important detritus dominated food chain, all illustrate that the Bao’an lake ecosystem is a mature and stable system according to the theories of Odum (1969). However, the moderate value of FCI along with extremely low values of CI and SOI still shows that the food web structure of the lake tends to lack complexity (Odum, 1969). Thus we conclude that the Bao'an lake ecosystem is a mature system but with a simple and vulnerable food web structure. This is mainly caused by unsustainable fish stocking and overfishing, since we know that prior to our study, piscivorious fish were targeted for removal from the lake, so that more traditional carp could be released. This resulted in significant changes in the fish community structure (Zhang et al., 1997; unpublished data). It has resulted in the mean trophic level of catch in Bao'an lake being only 2.26, far lower than Taihu lake (2.92) or Gehu lake (2.78), which indicates that Bao’an lake is suffering from overfishing and unsustainable fish stocking with too many low trophic level fish fry being released and caught in the lake leading to the simple food web structure.

Generally, this is the first ecosystem model of a shallow macrophytic lake in the middle reaches of the Yangtze River basin, which can quantitatively describe the food web structure and ecosystem properties. Based on the ecosystem analysis in Bao’an

Lake, we suggest that fish stocking in this lake should aim at increasing species richness, because numerous studies have showed that predator diversity can strengthen ecosystem function and food web structure (Griffin et al., 2008; Carey et al., 2011; Hargrave, 2009). We suggest that the lake be stocked with piscivorous and omnivorous species since these two groups seemed to be the key factors in mediating biodiversity–ecosystem functioning relationships (Petchey et al. 2004; Bruno et al., 2005). This fishery practice can also fully utilize the resources in this type of lake and increase the system’s SOI. This suggestion is also consistent with the former studies in these kinds of lakes (Xie et al., 2000; Cui and Li, 2005), which recommended that mandarin fish (*S. chuatsi*) and Chinese mitten crab (*Eriocheir sinensis*) can be reasonably stocked.

Additional attention should be given to the utilization of macrophytes. Some studies have indicated that dense vegetation could result in low feeding efficiency and poor growth of both small fish and piscivorous fishes (Colle and Shireman 1980; Miranda and Pugh 1997; Xie et al. 2005). Thus, herbivorous species with high economic value such as Chinese mitten crab (*Eriocheir sinensis*) and *Megalobrama amblycephala* Yih could be stocked moderately (Cui and Li, 2005).

5. General conclusions and perspective

Generally, the present study modeled the climate and anthropogenic effects on lake fish diversity and distribution patterns as well as the ecosystem structure and properties. Our results have drew the first global perspective on the lake fish distribution patterns, highlighted that climate variables have significant potential impacts on the fish diversity and distribution. We also examined the capacity of a new model approach MRT in predicting fish assemblages and species richness, defined the potential drive factors for the prediction. The results also highlighted the improvements of ensemble framework of SDM in modeling and predicting species distribution. Finally, a case study in a shallow lake also showed that anthropogenic activities could affect lake ecosystem function and structure, more precisely management activities should be taken to conserve and manage the lake ecosystem and aquatic resoureces.

However, further investigations are encouraged to go deeply on the mechanism of how fish diversity and distribution patterns response to climate change and human activities, more research also need to be taken to examine the interactions between global changes and anthropogenic effects. Further study of this approach using more accurate predictive models should be encouraged to advance our understanding of the profound influence of global change on species distribution under future scenarios.

While for the lake ecosystem structure, it should be acknowledged that the Ecopath model is just a steady-state model and therefore it cannot forecast the dynamic effects of fish stocking on the other organisms within the food web. The current study is just part of ongoing studies in these lakes, and further studies should be implemented to probe the carrying capacity of multi-species stocking based on the whole food web and ecosystem using the dynamic model (Ecosim). That will give more credible direction and prediction of the effects of fish stocking in this kind of lake.

References

1. Abell R, Thieme M L, Revenga C, Bryer M, Kottelat M, Bogutskaya N, Coad B, Mandrak N, Balderas S C, Bussing W, Stiassny M L J, Skelton P, Allen G R, Unmack P, Naseka A, Ng R, Sindorf N, Robertson J, Armijo E, Higgins J V, Heibel T J, Wikramanayake E, Olson D, López H L, Reis R E, Lundberg J G, Sabaj Pérez M H, Petry P. Freshwater ecoregions of the world: a new map of biogeographic units for freshwater biodiversity conservation. Bioscience, 2008, 58(5): 403-414
2. Aguilar-Kirigin Á J, Naya D E. Latitudinal patterns in phenotypic plasticity: the case of seasonal flexibility in lizards' fat body size. Oecologia, 2013, 173: 745-752
3. Allen K F. Relation between production and biomass. Journal of the Fisheries Research Board of Canada, 1971, 28(10): 1573-1581
4. Amarasinghe U S, Welcomme R L. An analysis of fish species richness in natural lakes. Environmental Biology of Fishes, 2002, 65(3): 327-339
5. Araújo M B, Guisan A. Five (or so) challenges for species distribution modelling. Journal of Biogeography, 2006, 33(10): 1677-1688
6. Araújo M B, Pearson R G, Thuiller W, Erhard M. Validation of species-climate impact models under climate change. Global Change Biology, 2005, 11(9): 1504-1513
7. Araújo M B, New M. Ensemble forecasting of species distributions. Trends in Ecology and Evolution, 2007, 22(1): 42-47
8. Araújo M B, Rahbek C. How does climate change affect biodiversity? Science-New York Then Washington, 2006, 313(5792): 1396
9. Barbet-Massin M, Walther B A, Thuiller W, Rahbek C, Jiguet F. Potential impacts of climate change on the winter distribution of Afro-Palaearctic migrant passerines. Biology Letters, 2009, 5(2): 248-251
10. Barradas J R S, Silva L G, Harvey B C, Fontoura N F. Estimating migratory fish

distribution from altitude and basin area: a case study in a large Neotropical river. Freshwater Biology, 2012, 57(11): 2297-2305

11. Beverton R J H, Holt S J. On the Dynamics of Exploited Fish Populations. Fishery investment, London, 1957
12. Biggerstaff B J. Comparing diagnostic tests: a simple graphic using likelihood ratios. Statistics in medicine, 2000, 19(5): 649-663
13. Boone R B, Krohn W B. Modeling the occurrence of bird species: are the errors predictable? Ecological Applications, 1999, 9(3): 835-848
14. Brazner J C, Beals E W. Patterns in fish assemblages from coastal wetland and beach habitats in Green Bay, Lake Michigan: a multivariate analysis of abiotic and biotic forcing factors. Canadian Journal of Fishery Aquatic Science, 1997, 54(8): 1743-1761.
15. Breiman L. Random forests. Machine Learning, 2001, 45(1): 5-32
16. Breiman L, Friedman J H, Olshen R A, Stone C G. Classification and regression trees. Wadsworth International Group, Belmont, California, USA, 1984
17. Breiman L. Bagging predictors. Machine Learning, 1996, 24(2): 123-140
18. Brotons L. Thuiller W, Araújo M B, Hirzel A H. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography, 2004, 27(4): 437-448
19. Brummer T J, Maxwell B D, Higgs M D, Rew L J. Implementing and interpreting local-scale invasive species distribution models. Diversity and Distributions, 2013, 19(8): 919-932
20. Bruno J F, O’Connor M I. Cascading effects of predator diversity and omnivory in a marine food web. Ecology Letters, 2005, 8(10): 1048-1056
21. Buisson L, Blanc L, Grenouillet G. Modelling stream fish species distribution in a river network: the relative effects of temperature versus physical factors. Ecology of Freshwater Fish, 2008a, 17(2): 244-257
22. Buisson L, Grenouillet G, Casajus N, Lek S. Predicting the potential impacts of climate change on stream fish assemblages. American Fisheries Society Symposium, 2010a, 73: 327-346

23. Buisson L, Grenouillet G. Contrasted impacts of climate change on stream fish assemblages along an environmental gradient. Diversity and Distributions, 2009, 15(4): 613-626

24. Buisson L, Thuiller W, Casajus N, Lek S, Grenouillet G. Uncertainty in ensemble forecasting of species distribution. Global Change Biology, 2010b, 16(4): 1145-1157

25. Buisson L, Thuiller W, Lek S, Lim P, Grenouillet G. Climate change hastens the turnover of stream fish assemblages. Global Change Biology, 2008b, 14(10): 2232-2248

26. Butchart S H W, Walpole M, Collen B, van Strien A, Scharlemann J P, Almond R E A, Baillie J E M, Bomhard B, Brown C, Bruno J, Carpenter K E, Carr G M, Chanson J, Chenery A M, Csirke J, Davidson N C, Dentener F, Foster M, Galli A, Galloway J N, Genovesi P, Gregory R D, Hockings M, Kapos V, Lamarque J, Leverington F, Loh J, McGeoch M A, McRae L, Minasyan A, Morcillo M H, Oldfield T E E, Pauly D, Quader S, Revenga C, Sauer J R, Skolnik B, Spear D, Stanwell-Smith D, Stuart S N, Symes A, Tierney M, Tyrrell T D, Vié J, Watson R. Global biodiversity: indicators of recent declines. Science, 2010, 328(5982): 1164-1168

27. Byron C, Link J, Costa-Piercea B, Bengtson D. Calculating ecological carrying capacity of shellfish aquaculture using mass-balance modeling: Narragansett Bay, Rhode Island. Ecological Modelling, 2011, 222(10): 1743-1755

28. Cao W.X., Zhang G.H., Ma J., Yu D.H., 1991. Preliminary studies on the phenomenon of size diminution of the fish resources in Lake Honghu. In: Honghu Research Group, Institute of Hydrobiology, CAS (Ed.) Studies on comprehensive exploitation of aquatic biological productivity and improvement of ecological environment in Lake Honghu. Beijing, China Ocean Press, pp.148–152 (in Chinese with English abstract).

29. Carey M P, Wahl D H. Fish diversity as a determinant of ecosystem properties across multiple trophic levels. Oikos, 2011, 120(1): 84-94

30. Cawsey E M, Austin M P, Baker B L. Regional vegetation mapping in Australia:

a case study in the practical use of statistical modelling. Biodiversity and Conservation, 2002, 11(12): 2239-2274

31. Chen, H. 1989. Impact of aquaculture on the ecosystem of the Donghu Lake, Wuhan. Acta Hydrobiologica Sinica 13(4), 359–368 (in Chinese with English abstract).
32. Cheng L, Lek S, Lek-Ang S, Li Z. Predicting fish assemblages and diversity in shallow lakes in the Yangtze River basin. Limnologica 2012, 42(2): 127-136
33. Cheung W W L, Lam V W Y, Sarmiento J L, Kearney K, Watson R, Pauly D. Projecting global marine biodiversity impacts under climate change scenarios. Fish and Fisheries, 2009, 10(3): 235-251
34. Christensen V, Walters C J. Ecopath with Ecosim: methods, capabilities, and limitation. Ecological Modelling, 2004, 172(2): 109-139
35. Christensen V. Ecosystem maturity-towards quantification. Ecological Modelling, 1995, 77(1): 3-32
36. Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 1960, 20(1): 37-46
37. Comte L, Buisson L, Daufresne M, Grenouillet G. Climate-induced changes in the distribution of freshwater fish: observed and predicted trends. Freshwater Biology, 2013, 58(4): 625-639
38. Crisp D T. Environmental requirements of common riverine European salmonid fish species in freshwater with particular reference to physical and chemical aspects. Hydrobiologia, 1996, 323(3): 201-221
39. Cucherousset J, Olden J D. Ecological impacts of nonnative freshwater fishes. Fisheries, 2011, 36(5): 215-230
40. Cui Y.B., Li Z.J., 2005. Fishery resources and conservation of environment in lakes of the Changjiang river basin. Beijing, Science Press.
41. Davidson T A, Sayer C D, Perrow M, Bramm M, Jeppesen E. The simultaneous inference of zooplanktivorous fish and macrophyte density from sub-fossil cladoceran assemblages: a multivariate regression tree approach. Freshwater Biology, 2010, 55(3): 546-564

42. De'ath G, Fabricius K E. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology, 2000, 81(11): 3178-3192

43. De'ath G. Multivariate regression trees: a new technique for modeling species-environment relationships. Ecology, 2002, 83(4): 1105-1117

44. DeFries R, Foley J, Asner G. Land-use choices: balancing human needs and ecosystem function. Frontiers in Ecology and the Environment, 2004, 2(5): 249-257

45. DeVantier L M, De’Ath G, Turak E, Done T J, Fabricius K E. Species richness and community structure of reef-building corals on the nearshore Great Barrier Reef. Coral Reefs, 2006, 25(3): 329-340

46. Dudgeon D, Arthington A H, Gessner M O, Kawabata Z I, Knowler D J, Leveque C, Naiman R J, Prieur-Richard A, Soto D, Stiassny M L J, Sullivan C A. Freshwater biodiversity: importance, threats, status and conservation challenges. Biological Reviews, 2006, 81(2): 163-182

47. Dufrene M, Legendre P. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecological Monograph, 1997, 67(3): 345-366

48. Elith J, Burgman M A. Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia, 2002

49. Elith J, Leathwick J R, Hastie T. A working guide to boosted regression trees. Journal of Animal Ecology, 2008, 77(4): 802-813

50. Elith J, Leathwick J R. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics, 2009, 40: 677-697

51. Elith J, Leathwick J. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. Diversity and Distributions, 2007, 13(3): 265-275

52. Fang J Y, Wang Z H, Zhao S Q, Li Y K, Tang Z Y, Yu D, Ni L Y, Liu H Z, Xie P, Da L J, Li Z Q, Zheng C Y. Biodiversity changes in the lakes of the Central Yangtze. Frontiers in Ecology and the Environment, 2006, 4(7): 369-377

53. Fetahia T, Schagerl M, Mengistoua S, Libralato S. Food web structure and trophic

interactions of the tropical highland lake Hayq, Ethiopia. Ecological Modelling, 2011, 222(3): 804-813

54. Fielding A H, Bell J F. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation, 1997, 24(1): 38-49
55. Finley J P. Tornado prediction. American Meteor Society, 1884
56. Finn J T. Measures of ecosystem structure and function derived from analysis of flows. Journal of Theoretical Biology, 1976, 56(2): 363-380
57. Friedman M. The Haredi (Ultra-Orthodox) Society—Sources, Trends and Processes. Jerusalem: The Jerusalem Institute for Israel Studies. 1991
58. Fu C Z, Wu J H, Chen J K, Wu Q H, Lei G C. Freshwater fish biodiversity in the Yangtze River basin of China: patterns, threats and conservation. Biodiversity and Conservation, 2003, 12(8): 1649-1685
59. Funtowicz S O, Ravetz J R. Uncertainty and Quality in Science for Policy. Kluwer Academic Publishers, Dortrecht, 1990
60. Gale M K, Hinch S G, Donaldson M R. The role of temperature in the capture and release of fish. Fish and Fisheries, 2013, 14(1): 1-33
61. Gevery M, Sans-Piche F, Grenouillet G, Tudesque L, Lek S. Modeling the impact of landscape types on the distribution of stream fish species. Canadian Journal of Fishery Aquatic Science, 2009, 66(3): 484-495
62. Gibson L A, Wilson B A, Cahill D M, Hill J. Spatial prediction of rufous bristlebird habitat in a coastal heathland: a GIS-based approach. Journal of Applied Ecology, 2004, 41(2): 213-223
63. Gillooly J F, Brown J H, West G B, Savage V M, Charnov E L. Effects of size and temperature on metabolic rate. Science, 2001, 293(5538): 2248-2251
64. Giraudoux P, Pleydell D, Raoul F, Quéré J P, Wang Q, Yang Y Y, Vuitton D A, Qiu J M, Yang W, Craigf P S. Transmission ecology of *Echinococcus multilocularis*: What are the ranges of parasite stability among various host communities in China? Parasitology international, 2006, 55: 237-246
65. Gislason H, Daan N, Rice J C, Pope J G. Size, growth, temperature and the

natural mortality of marine fish. Fish and Fisheries, 2010, 11(2): 149-158.

66. Grenouillet G, Buisson L, Casajus N, Lek S. Ensemble modelling of species distribution: the effects of geographical and environmental ranges. Ecography, 2011, 34(1): 9-17

67. Griffin J N, Haye K L, Hawkins S J, Thompson R C, Jenkins S R. Predator diversity and ecosystem function: density modifies the effect of resources partition. Ecology, 2008, 89(2): 298-305

68. Guan Z.H., 1995. Morphometric parameters of Bao’an lake and their limnological implications. In Resources, environment and fishery ecological management of macrophytic lakes. Science Press, Beijing (in Chinese with English abstract), pp. 3-15.

69. Guisan A, Graham C H, Elith J, Huettmann F. Sensitivity of predictive species distribution models to change in grain size. Diversity and Distribution, 2007, 13(3): 332-340

70. Guisan A, Thuiller W. Predicting species distribution: offering more than simple habitat models. Ecology Letter, 2005, 8(9): 993-1009

71. Guisan A, Zimmerman N E. Predictive habitat distribution models in ecology. Ecological Modelling, 2000, 135(2): 147-186

72. Guo C B, Ye S W, Lek S, Liu J S, Zhang T L, Li Z J. The need for improved fishery management in a shallow macrophytic lake in the Yangtze River basin: evidence from the food web structure and ecosystem analysis. Ecological Modelling, 2013, 267: 138-147

73. Gyllstrom M, Hansson L A, Jeppesen E, Garcia-Criado F, Gross E, Irvine K, Kairesalo T, Kornijow R, Miracle M R, Nykänen M, Noges T, Romo S, Stephen D, van Donk E, Moss B. The role of climate in shaping zooplankton communities of shallow lakes. Limnology and Oceanography, 2005, 50(6): 2008-2021

74. Hamann A, Gylander T, Chen P Y. Developing seed zones and transfer guidelines with multivariate regression trees. Tree Genetics & Genomes, 2011, 7(2): 399-408

75. Hargrave C W. Effects of fish species richness and assemblage composition on

stream ecosystem function. Ecology of Freshwater Fish, 2009, 18(1): 24-32

76. Hastie T J, Tibshirani R J. Generalized Additive Models (No. 43). CRC Press, 1990

77. Hawkins B A, Field R, Cornell H V, Currie D J, Guégan J, Kaufman D M, Kerr J T, Mittelbach G G, Oberdorff T, O'Brien E M, Porter E E, Turner J R G. Energy, water, and broad-scale geographic patterns of species richness. Ecology, 2003, 84(12): 3105-3117

78. He Y F, Wang J W, Lek S, Cao W X, Lek-ang S. Structure of endemic fish assemblages in the upper Yangtze River basin. River research and application, 2011, 27(1): 59-75

79. He Y, Wang J, Lek-Ang S, Lek S. Predicting assemblages and species richness of endemic fish in the upper Yangtze River. Science of the Total Environment, 2010, 408(19): 4211-4220

80. Heggenes J, Krog O M W, Lindas O R, Dokk J G, Bremnes T. Homeostatic behavioural responses in a changing environment: brown trout (*Salmo trutta*) become nocturnal during winter. Journal of Animal Ecology, 1993, 62(2): 295-308

81. Heikkinen R K, Luoto M, Araujo M B, Virkkala R, Thuiller W, Sykes M T. Methods and uncertainties in bioclimatic envelope modelling under climate change. Progress in Physical Geography, 2006, 30(6): 751-777

82. Hernandez P A, Graham C H, Master L L, Albert D L. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography, 2006, 29(5): 773-785

83. Hijmans R J, Cameron S E, Parra J L, Jones P G, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 2005, 25(15): 1965-1978

84. Hijmans R J, Cameron S E, Parra J L, Jones P G, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 2005, 25(15): 1965-1978

85. Hossaina M M, Matsuishi T, Arhonditsis G. Elucidation of ecosystem attributes

of an oligotrophic lake in Hokkaido, Japan, using Ecopath with Ecosim (EwE). Ecological Modelling, 2010, 221(13): 1717-1730

86. Hu C.L, Huang X.F, 1991. Collected papers on the fishery ecology and exploition technology of the Bao'an lake. Science Press, Beijing(in Chinese with English abstract).

87. Hutchinson M F. Interpolating mean rainfall using thin plate smoothing splines. International Journal of geographical information systems, 1995, 9(4): 385-403

88. IPCC (Intergovernmental Panel on Climate Change). The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In: Stocker T F, Qin D, Plattner G, Tignor M, Allen S K, Boschung J, Nauels A, Xia Y, Bex V, Midgley P M (eds.). Cambridge and New York: Cambridge University Press, 2013

89. Jeppesen E, Kronvang B, Meerhoff M, Søndergaard M, Hansen K M, Andersen H E, Lauridsen T L, Beklioglu M, Ozen A, Olesen J E. Climate change effects on runoff, catchment phosphorus loading and lake ecological state, and potential adaptations. Journal of Environmental Quality, 2009, 38(5): 1930-1941

90. Jeppesen E, Meerhoff M, Holmgren K, Gonzalez-Bergonzoni I, Teixeira-De Mello F, Declerck S A J, De Meester L, Søndergaard M, Lauridsen T L, Bjerring R, Conde-Porcuna J M, Mazzeo N, Iglesias C, Reizenstein M, Malmquist H J, Liu Z W, Balayla D, Lazzaro X. Impacts of climate warming on lake fish community structure and potential effects on ecosystem function. Hydrobiologia, 2010, 646: 73-90

91. Jeppesen E, Søndergaard M, Meerhoff M, Lauridsen T L, Jensen J P. Shallow lake restoration by nutrient loading reduction-some recent findings and challenges ahead. Hydrobiologia, 2007, 584(1): 239-252

92. Jia P Q, Hu M H, Hu Z J, Liu Q G, Wu Z. Modeling trophic structure and energy flows in a typical macrophyte dominated shallow lake using the mass balanced model. Ecological Modelling, 2012, 233: 26-30

93. Jiménez-Valverde A. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution

modelling. Globle Ecology Biogeography, 2012, 21(4): 498-507

94. Jin G., 2001. Studies on the biology and fishery management of Chinese Mitten Crab, Ericheir sinensis. PhD Dissertation. Wuhan: Institute of Hydrobiology, Chinese Academy of Sciences (in Chinese with English abstract).

95. Jin, X.C. 1995. Environment of Chinese lakes, volumes I, II and II. Ocean Press, Beijing (In Chinese).

96. Johnson J A, Kelsch S W. Effects of evolutionary thermal environment on temperature-preference relationships in fishes. Environtal Biology of Fish, 1998, 53(4): 447-458

97. Joy M K, Death R G. Predictive modelling and spatial mapping of freshwater fish and decapod assemblages using GIS and neural networks. Freshwater Biology, 2004, 49(8): 1036-1052

98. Kadmon R., Farber O, Danin A. A systematic analysis of factors affecting the performance of climatic envelope models. Ecological Application, 2003, 13(3): 853-867

99. Kang B, Deng J, Wu Y, Chen L, Zhang J, Qiu H, Lu Y, He D. Mapping China's freshwater fishes: diversity and biogeography. Fish and Fisheries, 2013, doi: 10.1111/faf.12011

100. Landis J R, Koch G G. The measurement of observer agreement for categorical data. Biometrics, 1977, 33(1): 159-174

101. Larsen D R, Speckman P L. Multivariate regression trees for analysis of abundance data. Biometrics, 2004, 60(2): 543-549

102. Lauzeral C, Leprieur F, Beauchard O, Duron Q, Oberdorff T, Brosse S. Identifying climatic niche shifts using coarse-grained occurrence data: a test with non-native freshwater fish. Global Ecology and Biogeography, 2011, 20(3): 407-414

103. Lawler J J, White D, Neilson R P, Blaustein A R. Predicting climate-induced range shifts: model differences and model reliability. Global Change Biology, 2006, 12(8): 1568-1584

104. Lazzaro X. Do the trophic cascade hypothesis and classical biomanipulation

approaches apply to tropical lakes and reservoirs? Verhandlungen der Internationale Vereinigung fur Limnologie, 1997, 26: 719-730

105. Legendre P. Spatial autocorrelation: trouble or new paradigm? Ecology, 1993, 74(6): 1659-1673

106. Levinsky I, Skov F, Svenning J C, Rahbek C. Potential impacts of climate change on the distributions and diversity patterns of European mammals. Biodiversity and Conservation, 2007, 16(13): 3803-3816

107. Li S.Z. (1981) Studies on Zoogeographical Divisions for Freshwater Fishes in China. Science Press, Beijing (in Chinese).

108. Li W., Zhang T.L., Li Z.J. 2010. Spatial distribution and abundance of small fishes in Xiaosihai Lake, a shallow lake along the Changjiang (Yangtze) River, China. Chinese Journal of Oceanology and Limnology 28 (3),470-477.

109. Li Y K, Chen Y, Song B, Derek O, Yu N, Chen L Q. Ecosystem structure and functioning of Lake Taihu (China) and the impacts of fishing. Fisheries Research 2009, 95: 309-324

110. Liang Y.L. and Liu H.Q. (1995) Resources, environment and fishery ecological management of macrophytic lakes. The Science Press, Beijing (in Chinese with English abstract).

111. Libralato S, Coll M, Tudela S, Palomera I, Pranovi F. Novel index for quantification of ecosystem effects of fishing as removal of secondary production. Marine Ecology Progress Series, 2008, 355: 107-129

112. Lindeman RL. The trophic-dynamic aspect of ecology. Ecology, 1942, 23(4): 399-417

113. Liu C, Berry P M, Dawson T P, Pearson R G. Selecting thresholds of occurrence in the prediction of species distributions. Ecography, 2005, 28(3): 385-393

114. Liu C, White M, Newell G. Measuring and comparing the accuracy of species distribution models with presence-absence data. Ecography, 2011, 34(2): 232-243

115. Liu H.Q., Li Z.J., Fang R.L., 1995. A series of studies on high quality and high efficiency ecological fishery model of Bao’an Lake. In Resources, environment and fishery ecological management of macrophytic lakes. The Science Press,

Beijing (in Chinese with English abstract), p.228-236.

116. Liu J K. Lakes of the middle and lower basins of the Changjiang (China). In: Taub F B (eds.), Ecosystems of the world 23 Lakes and reservoirs. pp. 331-355. Amsterdam: Elsevier, 1984

117. Liu Q G, Chen Y, Li J L, Chen L Q. The food web structure and ecosystem properties of a filter-feeding carps dominated deep reservoir ecosystem. Ecological Modelling, 2007, 203(3): 279-289

118. Liu X, Wang H. Estimation of minimum area requirement of river-connected lakes for fish diversity conservation in the Yangtze River floodplain. Diversity and Distributions, 2010, 16(6): 932-940

119. Liu, J., He, B. 1992. Cultivation of the Chinese freshwater fishes, The Science Press, Beijing.pp. 381 (in Chinese with English abstract).

120. Liu, J.K., 1999. Advanced Aquatic Biology. Science Press, Beijing.

121. Lobo J M, Jiménez-Valverde A, Real R. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography, 2008, 17(2): 145-151

122. Luck G W. The habitat requirements of the rufous treecreeper (*Climacteris rufa*). 1. Preferential habitat use demonstrated at multiple spatial scales. Biological Conservation, 2002, 105(3): 383-394

123. Luoto M, Heikkinen R K, Pöyr J, Saarinen K. Determinants of the biogeographical distribution of butterflies in boreal regions. Journal of Biogeography, 2006, 33(10): 1764-1778

124. Luoto M, Pöyry J, Heikkinen R K, Saarinen K. Uncertainty of bioclimate envelope models based on the geographical distribution of species. Global Ecology and Biogeography, 2005, 14(6): 575-584

125. Magurran A E. Threats to freshwater fish. Science, 2009, 325(5945): 1215-1216

126. Maloney K O, Weller D E, Michaelson D E, Ciccotto P J. Species distribution models of freshwater stream fishes in Maryland and their implications for management. Environmental Modeling & Assessment, 2013, 18(1): 1-12

127. Manel S, Dias J M, Ormerod S J. Comparing discriminant analysis, neural

networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. Ecological Modelling, 1999, 120(2): 337-347

128. Manel S, Williams H C, Ormerod S J. Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology, 2001, 38(5): 921-931

129. Mann R H K. Environmental requirements of European non-salmonid fish in rivers. Hydrobiologia, 1996, 323(3): 223-235

130. Marmion M, Luoto M, Heikkinen R K, Thuiller W. The performance of state-of-the-art modelling techniques depends on geographical distribution of species. Ecological Modelling, 2009a, 220(24): 3512-3520

131. Marmion M, Parviainen M, Luoto M, Heikkinen R K, Thuiller W.Evaluation of consensus methods in predictive species distribution modelling. Diversity and Distribution, 2009b, 15(1): 59-69

132. McCullagh P, Nelder J A, Generalized linear models (Monographs on statistics and applied probability 37). London: Chapman Hall, 1989

133. McCullagh P, Nelder J A. Generalized linear models. London: Chapman and Hall, 1983

134. McPherson J, Jetz W. Effects of species’ ecology on the accuracy of distribution models. Ecography, 2007, 30(1): 135-151

135. Mello T D, Meerhoff M, Pekcan-Hekim Z, Jeppesen E. Substantial differences in littoral fish community structure and dynamics in subtropical and temperate shallow lakes. Freshwater Biology, 2009, 54(6): 1202-1215

136. Mika A M, Weiss R M, Olfert O, Hallett R H, Newman J A. Will climate change be beneficial or detrimental to the invasive Swede midge in North America? Contrasting predictions using climate projections from different general circulation models. Global Change Biology, 2008, 14(8): 1721-1733

137. Millenium Ecosystem Assessment, Ecosystems and Human Well-being: Biodiversity Synthesis. World Resources Institute, Washington D.C., USA, 2005

138. Morissette L, Hammill M O, Savenkoff C. The trophic role of marine mammals in the northern Gulf of St Lawrence. Marine Mammal Science, 2006, 22(1):

74-103

139. Odum E P. Fundamental of ecology. Philadelphia: Saunders, 1971

140. Odum E P. The strategy of ecosystem development. Sustainability, 1969, 164: 58

141. Odum W E, Heald E J. The detritus-based food web of an estuarine mangrove community. In: Cronin L E (ed.). pp. 265-286. New York: Academic Press, 1975

142. OhlemÜller R, Gritti E S, Sykes M T, Thomas C D. Quantifying components of risk for European woody species under climate change. Global Change Biology, 2006, 12(9): 1788-1799

143. Olden J D, Rooney T P. On defining and quantifying biotic homogenization. Global Ecology and Biogeography, 2006, 15(2): 113-120

144. Ouellette M H, Legendre P, Borcard D. Cascade multivariate regression tree: a novel approach for modelling nested explanatory sets. Methods in Ecology and Evolution, 2012, 3(2): 234-244

145. Park Y S, Grenouillet G, Esperance B, Lek S. Stream fish assemblages and basin land cover in a river network. Science of the Total Environment, 2006, 365(1): 140-153

146. Ouellette M H, Legendre P, Borcard D. Cascade multivariate regression tree: a novel approach for modelling nested explanatory sets. Methods in Ecology and Evolution, 2012, 3(2): 234-244

147. Palomares M L D, Pauly D. Predicting food consumption of fish populations as functions of mortality, food type, morphometrics, temperature and salinity. Marine and Freshwater Research, 1998, 49(5): 447-453

148. Park Y S, Grenouillet G, Esperance B, Lek S. Stream fish assemblages and basin land cover in a river network. Science of the Total Environment, 2006, 365(1): 140-153

149. Park R A. A generalized model for simulating lake ecosystems. Simulation, 1974, 23: 33-50

150. Pauly D, Christensen V, Walters C. Ecopath, Ecosim, and Ecospace as tools for evaluating ecosystem impact of fisheries. Journal of Marine Science, 2000, 57(3): 697-706

151. Pauly D, Christensen V. Primary production required to sustain global fisheries. Nature, 1995, 374(6519): 255-257

152. Pearce J, Ferrier S. Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling, 2000, 133(3): 225-245

153. Pearson R G, Dawson T P. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography, 2003, 12(5): 361-371

154. Pearson R G, Thuiller W, Araújo M B, Martinez-Meyer E, Brotons L, McClean C, Miles L, Segurado P, Dawson T P, Lees D C. Model-based uncertainty in species range prediction. Journal of Biogeography, 2006, 33(10): 1704-1711

155. Pearson R G. Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org, 2007

156. Pearson R G. Species' Distribution Modeling for Conservation Educators and Practitioners. Synthesis. American Museum of Natural History, 2007. Available at: http://ncep.amnh.org

157. Petchey O L, Downing A L, Mittelbach G G, Persson L,Steiner C F, Warren P H, Woodward G. Species loss and the structure and functioning of multitrophic aquatic systems. Oikos, 2004, 104(3): 467-478

158. Peterson A T, Papes M, Soberón J. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecological Modelling, 2008, 213(1): 63-72

159. Peterson A T. Predicting the geography of species' invasions via ecological niche modeling. The quarterly review of Biology, 2003, 78(4): 419-433

160. Peterson B J, Holmes R M, McClelland J W, Vorosmarty C J, Lammers R B, Shiklomanov A I, Shiklomanov I A, Rahmstorf S. Increasing river discharge to the Arctic Ocean. Science, 2002, 298(5601): 2171-2173

161. Petry P, Bayley P, Markle D. Relationships between fish assemblages, macrophytes and environmental gradients in the Amazon River floodplain. Journal of Fish Biology, 2003, 63(3): 547-579

162. Phillips S J, Anderson R P, Schapire R E. Maximum entropy modeling of species geographic distributions. Ecological Modelling, 2006, 190(3): 231-259

163. Pont D, Hugueny B, Oberdorff T. Modelling habitat requirement of European fishes: do species have similar responses to local and regional constraints? Canadian Journal of Fisheries and Aquatic Sciences, 2005, 62(1): 163-173

164. Pottier J, Dubuis A, Pellissier L, Maiorano L, Rossier L, Randin C F, Vittoz P, Guisan A. The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. Global Ecology and Biogeography, 2013, 22(1): 52-63

165. Prasad A K, Singh R P, Kafatos M. Influence of coal based thermal power plants on aerosol optical properties in the Indo-Gangetic basin. Geophyscience Research Letter, 2006, 33(5): 1-4

166. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org/.2013

167. Rathert D, White D, Sifneos J C, Hughes R M. Environmental correlates of species richness for native freshwater fish in Oregon, USA. Journal of Biogeography, 1999, 26(2): 257-273

168. Reiss H, Cunze S, Koenig K, Neumann H, Kroencke I. Species distribution modelling of marine benthos: a North Sea case study. Marine Ecology Progress Series, 2011, 442: 71-86

169. Ren M.L. (1994) Fish fauna of Heilongjiang. Chinese Journal of Fisheries,1:1-14.

170. Ripley B D. Pattern recognition and neural networks. Cambridge: Cambridge University Press, 1996

171. Roura-Pascual N, Brotons L, Peterson A T, Thuiller W. Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. Biological Invasions, 2009, 11(4): 1017-1031

172. Rubidge E M, Monanan W B, Parra J L, Cameron S E, Brashares J S. The role of climate, habitat, and species cooccurrence as drivers of change in small mammal distributions over the past century. Global Change Biology, 2011, 17(2): 696-708

173. Rushton S P, Ormerod S J, Kerby G. New paradigms for modelling species distributions? Journal of applied ecology, 2004, 41(2): 193-200

174. Scavia D, Bloomfield J A, Fisher J S. Documentation of CLEANX: a generalized model for simulating the open-water ecosystems of lakes. Simulation, 1974, 23(2): 51-56

175. Segurado P, Araujo M B. An evaluation of methods for modelling species distributions. Journal of Biogeography, 2004, 31(10): 1555-1568

176. Sharma S, Jackson D A, Minns C K, Shuter B J. Will northern fish populations be in hot water because of climate change? Global Change Biology, 2007, 13(10): 2052-2064

177. Shi C.Y. (1989) A general outline of Chinese lakes. Science Press, Beijing (in Chinese).

178. Stanley E, Doyle M W. Trading off: the ecological effects of dam removal. Frontiers of Ecology Environment, 2003, 1(1): 15-22

179. Stockwell D R, Peterson A T. Effects of sample size on accuracy of species distribution models. Ecological Modelling, 2002, 148(1): 1-13

180. Stojkovic M, Simic V, Milosevic D, Mancev D, Penczak T. Visualization of fish community distribution patterns using the self-organizing map: A case study of the Great Morava River system (Serbia). Ecological Modelling, 2013, 248: 20-29

181. Su Z.G., Zhang T.L., Cai Q.H., 1995. On change of aquatic vegetation in Bao’an Lake with remarks on its fishery uyilization. In Resources, environment and fishery ecological management of macrophytic lakes. The Science Press, Beijing (in Chinese), p. 147-159.

182. Swets J A. Measuring the accuracy of diagnostic systems. Science, 1988, 240(4857): 1285-1293

183. Taniguchi Y, Rahel F J. Novingen D C, Gerow K G. Temperature mediation of competitive interactions among three fish species that replace each other along longitudinal stream gradients. Canadian Journal of Fisheries and Aquatic Sciences, 1998, 55(8) 1894-1901

184. Tejerina-Garro F, Fortin R, Rodriguez M. Fish community structure in relation to

environmental variation in floodplain lakes of the Araguaia River, Amazon Basin. Environment Biology of Fishes, 1998, 51(4): 399-410

185. Thuiller W, Broennimann O, Hughes G, Alkemande J R M, Midgley G F, Corsi F. Vulnerability of African mammals to anthropogenic climate change under conservative land transformation assumptions. Global Change Biology, 2006, 12(3): 424-440

186. Thuiller W, Lavorel S, Araújo M B, Sykes M T, Prentice I C. Climate change threats to plant diversity in Europe. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(23): 8245-8250

187. Thuiller W. BIOMOD–optimizing predictions of species distributions and projecting potential future shifts under global change. Global Change Biology, 2003, 9(10): 1353-1362

188. Thuiller W. Patterns and uncertainties of species' range shifts under climate change. Global Change Biology, 2004, 10(12): 2020-2027

189. Tisseuil C, Leprieur F, Grenouillet G, Vrac M, Lek S. Projected impacts of climate change on spatio-temporal patterns of freshwater fish beta diversity: a deconstructing approach. Global Ecology and Biogeography, 2012, 21(12): 1213-1222

190. Tong L.,1999. Ecopath model-A mass-balance modeling for ecosystem estimation.20 (2),103-107 (in Chinese with English abstract).

191. Tsoar A, Allouche O, Steinitz O, Rotem D, Kadmon R. A comparative evaluation of presence-only methods for modelling species distribution. Diversity and Distribution, 2007, 13(4): 397-405

192. Ulanowicz R E, Puccia C J. Mixed trophic impacts in ecosystem. Coenoses, 1990, 5(1): 7-16

193. Ulanowicz R E. Ecosystem Trophic Foundations: Lindeman Exonerata. In: Patten B C, Jørgensen S E (eds.). Complex ecology: the part-whole relation in ecosystems. pp. 549-560. Englewood Cliffs, Prentice Hall, 1995

194. Ulanowicz R E. Growth and development: ecosystem phenomenology. New York: Springer Verlag, 1986

195. Vayssières M P, Plant R E, Allen-Diaz B H. Classification trees: an alternative non-parametric approach for predicting species distributions. Journal of Vegetation Science, 2000, 11(5): 679-694

196. Villanueva M C, Isumbisho M, Kaningini B, Moreau J, Micha J C. Modeling trophic interactions in Lake Kivu: what roles do exotics play? Ecological Modelling, 2008, 212(3): 422-438

197. Virkkala R, Heikkinen R K, Leikola N, Luoto M. Projected large-scale range reductions of northern-boreal land bird species due to climate change. Biological Conservation, 2008, 141(5): 1343-1353

198. Virkkala R, Luoto M, Heikkinen R K, Leikola N. Distribution patterns of boreal marshland birds: modelling the relationships to land cover and climate. Journal of Biogeography, 2005, 32(11): 1957-1970

199. Vörösmarty C J, McIntyre P B, Gessner M O, Dudgeon D, Prusevich A, Green P, Glidden S S, Bunn E, Sullivan C A, Reidy Liermann C, Davies P M. Global threats to human water security and river biodiversity, Nature, 2010, 467(7315): 555-561

200. Wang J., Liang Y.L.,1995. A preliminary study on the structure and abundance of the attached algae in Bao’an lake. In Resources, environment and fishery ecological management of macrophytic lakes. The Science Press, Beijing (in Chinese), pp. 228-236

201. Wang S.M., Dou H.S. (1998) Lakes of China. Science Press, Beijing.

202. Wetzel R G. Limnology: lake and river ecosystems. Academic press, 2001

203. Whittaker R J, Nogués-Bravo D, Araújo M B. Geographical gradients of species richness: a test of the water-energy conjecture of Hawkins et al. (2003) using European data for five taxa. Global Ecology and Biogeography, 2007, 16(1): 76-89

204. Williams P, Whitfield M, Biggs J, Bray S, Fox G, Nicolet P, Sear D. Comparative biodiversity of rivers, streams, ditches and ponds in an agricultural landscape in Southern England. Biological Conservation, 2004, 115(2): 329-341

205. Wolter C. Temperature influence on the fish assemblage structure in a large

lowland river, the lower Oder River, Germany. Ecology of Freshwater Fish, 2007, 16(4): 493-503

206. Wright D H. Species-energy theory: an extension of species area theory. Oikos, 1983, 41: 496-506

207. Wu J G, Huang J H, Han X G, Gao X M, He F L, Jiang M X, Jiang Z G, Primack R B, Shen Z H. The three gorges dam: an ecological perspective. Frontier of Ecology and Environment, 2004, 2(5): 241-248

208. Wu X.W. (1964) The Cyprinoid Fishes of China, Vol. 1. People's Press, Beijing (in Chinese).

209. Wu X.W. (1977) The Cyprinoid Fishes of China, Vol. 2. People's Press, Beijing (in Chinese).

210. Xie S G, Cui Y B, Li Z J. Small fish communities in two regions of the Liangzihu Lake, China, with or without submersed macrophytes. Journal of Applied Ichthyology, 2001, 17: 89-92

211. Xie S G, Cui Y B, Zhang T L, Fang R L, Li Z J. The spatial pattern of the small fish community in the Biandantang Lake-a small shallow lake along the middle reach of the Yangtze River, China. Environmental Biology of Fishes, 2000, 57(2): 179-190

212. Xie S G, Cui Y B. The spatial pattern of the small fish community in the Biandantang Lake – a small shallow lake along the middle reach of the Yangtze River, China. Environment Biology of Fish, 2000, 57: 179-190

213. Xie P, Chen Y. Threats to biodiversity in Chinese inland waters. Ambio, 1999, 28: 674-681

214. Xie P. Three-Gorges Dam: risk to ancient fish. Science, 2003, 302(5648): 1149-1151

215. Xu S N, Chen Z, Li C, Huang X P, Li S Y. Assessing the carrying capacity of tilapia in an intertidal mangrove-based polyculture system of Pearl River Delta, China. Ecological Modelling, 2011a, 222(3): 846-856

216. Xu S N, Chen Z, Li S, He P. Modeling trophic structure and energy flows in a coastal artificial ecosystem using mass-balance Ecopath model. Estuaries and

Coasts, 2011b, 34(2): 351-363

217. Yan Y, Liang Y. Energyflow of macrozoobenthic community in a macrophytic lake, Biandantang Lake. Acta Ecologica Sinica 2003, 23 (3): 527–538 (in Chinese with English abstract)

218. Yang F，Lu X，Lou Y，Xue B，Yao S，Wei X，Li Z, Shan Y. Structures of fish assemblages from lakes in Songnen Plain. Journal of Lake Science, 2010,22: 842-851(in Chinese with English abstract)

219. Yang Y M, Tian K, Hao J M, Pei S J, Yang Y X. Biodiversity and biodiversity conservation in Yunnan, China. Biodiversity and Conservation, 2004, 13(4): 813-823

220. Ye S W, Li Z, Lek-Ang S, Feng G, Lek S, Cao W. Community structure of small fishes in a shallow macrophytic lake (Niushanhu Lake) along the middle reach of the Yangtze River, China. Aquatic Living Resources, 2006, 19: 349-359

221. Yin M. Ecology of fish [M].China agriculture Press, Beijing, 1982 (In Chinese)

222. Yoon J D, Kim J H, Byeon M S, Yang H J, Park J Y, Shim J H, Song H B, Yang H, Jang M H. Distribution patterns of fish communities with respect to environmental gradients in Korean streams. Annales de Limnologie - International Journal of Limnology, 2011, 47: 63-71

223. Yoon J D, Kim J H, Byeon M., Yang H J, Park J Y, et al. Distribution patterns of fish communities with respect to environmental gradients in Korean streams. Annales de Limnologie International Journal of Liminology, 2011, 47(1): 63-71

224. Yu D, Chen M, Zhou Z C, Eric R, Tang Q Y, Liu H Z. Global climate change will severely decrease potential distribution of the East Asian coldwater fish *Rhynchocypris oxycephalus* (Actinopterygii, Cyprinidae). Hydrobiologia, 2013, 700(1): 23-32

225. Zhang CL. Distribution of Chinese freshwater fishes. Acta Geographica Sinica 1954,3:279-284(in Chinese)

226. Zhang R.Z. Zoogeography of China. Science Press, 1999, Beijing (in Chinese).

227. Zhang TL. Life-history strategies, trophic patterns and community structure in the fishes of Lake Biandantang. PhD Dissertation. Institute of Hydrobiology, Chinese

Academy of Sciences, Wuhan (in Chinese with English abstract), 2005

228. Zhao S, Fang J, Peng C, Tang Z, Piao S. Patterns of fish species richness in China's lakes. Global Ecology and Biogeography, 2006, 15(4): 386-394

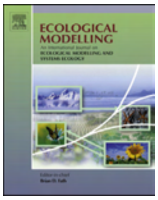
Part II: Publications

Ecological Modelling 267 (2013) 138–147

Contents lists available at ScienceDirect

Ecological Modelling

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



The need for improved fishery management in a shallow macrophytic lake in the Yangtze River basin: Evidence from the food web structure and ecosystem analysis

Chuanbo Guo[a,b,c], Shaowen Ye[a,*], Sovan Lek[c], Jiashou Liu[a], Tanglin Zhang[a], Jin Yuan[a], Zhongjie Li[a,*]

[a] Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China
[b] University of the Chinese Academy of Sciences, Beijing 100039, China
[c] Université de Toulouse III, UMR 5174 EDB, CNRS, 118 Route de Narbonne, 31062 Toulouse Cedex 09, France

ARTICLE INFO

Article history:
Received 19 June 2013
Received in revised form 8 July 2013
Accepted 12 July 2013
Available online 1 September 2013

Keywords:
Food web
Ecosystem properties
Macrophytic lake
Yangtze River
Ecopath with Ecosim
Ecosystem-based fishery management

ABSTRACT

There are numerous shallow macrophytic lakes distributed in the middle reaches of the Yangtze River basin, which are an important fishery resource for this part of China. However, there is limited knowledge about the food web structure and energy flows of these highly disturbed ecosystems, mainly due to lack of suitable ecosystem approaches applied to the abundant but isolated ecological data from these lakes. To better manage the important ecosystems, Ecopath with Ecosim was applied to establish a mass-balance model for a typical shallow macrophytic lake (Bao'an Lake) as a case study, with the aim of describing the food web structure and the properties of the ecosystem to evaluate the ecological implications for fishery resource management and the protection of the aquatic ecosystem of these lakes. Given that there were extensive first-hand data available for the target lake, a credible trophic model including 23 functional groups was constructed. The results showed that all the commercial fish groups suffered from high fishing pressure for their higher ecotrophic efficiency (EE) values, normally more than 0.5. On the contrary, forage resources such as attached algae, submerged plants and molluscs were not fully utilized by the lake fishery, with EE values even as low as 0.089, 0.120 and 0.126 respectively. The discrete trophic level of large culters was highest (3.143) in the lake ecosystem, followed by mandarin fish (3.138) and snakehead fish (3.131). For the transfer efficiencies in the food web structure, a mean value of 8.68% was calculated for the lake ecosystem. Ecosystem maturity indices such as TPP/TR (1.640), TPP/TB (6.993), as well as ascendency (0.387) which were derived from the network analysis together with the revealed detritus-based trophic flow, illustrated that the Bao'an Lake ecosystem was a mature system according to Odum's theory. However when compared with some other lake ecosystems, the Bao'an Lake ecosystem, as well as some China lake ecosystems, showed extremely low values of CI (Connectance index), FCI (Finn's cycling index) and SOI (system omnivory index), indicating that the food web structures of these Chinese lake ecosystems tended to be simpler and more linear than lake ecosystems in other countries. Consequently, this study established the first food web model for a shallow macrophytic lake and provided overall insights and ecosystem knowledge for this kind of shallow macrophytic lake, and indicated an urgent need for fishery resources management to shift from traditional population-based to ecosystem-based models.

1. Introduction

There are numerous freshwater lakes distributed in the middle reaches of the Yangtze River basin (MYRB), which cumulatively represents about 30% of the total lake area in China (Liu and He, 1992). These subtropic lakes are generally shallow and highly productive, with a diverse fish community (40–70 species or even more in some lakes), abundant invertebrate resources and dense submerged macrophytes (Liu, 1984; Xie and Chen, 1996), thus representing one of the most important fishery bases in China (Cui and Li, 2005). Due to their significant ecological value, submerged macrophytes in shallow lakes are usually known as "underwater forest", and they provide high quality food as well as appropriate habitats for aquatic animals (Hu and Huang, 1991). However, local fisheries have heavily exploited living aquatic resources in these shallow macrophytic lakes during the past few decades.

---

* Corresponding authors. Tel.: +86 27 68 78 00 63; fax: +86 27 68 78 00 63.
E-mail addresses: yesw@ihb.ac.cn (S. Ye), zhongjie@ihb.ac.cn (Z. Li).

0304-3800/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecolmodel.2013.07.013

Specifically, overstocking of grass carp (*Ctenopharyngodon idellus*), bighead carp (*Aristichthys nobilis*), silver carp (*Hypophthalmichthys molitrix*) and common carp (*Cyprinus carpio*) was common in these lakes (Liu and He, 1992). Such fishery practices often caused water quality problems and even ecological deterioration, as overstocking of grass carp can result in a drastic reduction or total elimination of submerged macrophytes and a subsequent increase in algal biomass. To obtain higher yields of bighead and silver carp in the lakes, sewage and fertilizers have often been introduced to increase plankton production (Chen, 1989). Additionally, the economic benefit associated with the fisheries dropped sharply year after year as a result of unsuitable management (Hu and Huang, 1991).

In order to reconcile fishery development with environmental protection in the lakes, the key question is to determine how to transform the traditional fishery management approaches to use the resources more rationally (Hu and Huang, 1991; Liang and Liu, 1995; Cui and Li, 2005). As part of this objective, two key research projects focusing on lake fishery resources and environmental management were launched in the 1990s by the Ministry of Agriculture of China in two successive 5-year periods, covering the period of "the seventh five-year plan" and "the eighth five-year plan". Bao'an Lake, a typical shallow macrophtyic lake in the MYRB, was included as an important experimental and demonstration lake in the two projects. Thus extensive first-hand data on the Bao'an Lake ecosystem was collected during that period (Hu and Huang, 1991; Liang and Liu, 1995). However, due to the limited use of ecosystem approaches, most of those studies focused on descriptive studies of individual organisms, environment, or rough assessments of single species stocking capacity, mostly traditional Chinese carp, and thus lacked an analysis of the interactions between different components in the ecosystem. Therefore conclusions may be of limited value since food web structure and trophic interactions between major components within an ecosystem are extraordinarily complicated (Odum, 1969). Quantitative studies on the food web structure, energy flows and ecosystem properties of Bao'an Lake are therefore urgently required for fishery development and model transformation.

Ecopath with Ecosim (EwE) (Christensen et al., 2005) has been widely considered to be an appropriate tool for the analysis of food webs. The Ecopath model is ecosystem-based software, and like other ecosystem models, can summarize and view changes in species interactions within an ecosystem (Christensen et al., 2005; Xu et al., 2011b). Both direct and indirect effects of species changes in the ecosystem can be explored and effects on the overall functioning of the ecosystem can be estimated (Heymans et al., 2004). Furthermore, the suitability of this approach is its application to a broad field of theories that are useful for ecosystem studies, e.g. thermodynamic concepts, information theory, trophic level description and network analysis (Xu et al., 2011a). EwE was first introduced to China by Tong (1999) and thereafter has been used in several aquatic ecosystems to manage fisheries in China, especially in ocean systems. However, few EwE models have been constructed for China's lake ecosystems except Li et al. (2009) and Jia et al. (2012). So far, no EwE model is available for the shallow macrophytic lakes in the MYRB, which have long been used for stocking fish. These lakes have unique aquatic living resources and environmental conditions.

Therefore, currently, a EwE model of Bao'an Lake was constructed using the extensive data already collected. This study will be significant not only for environment and ecosystem conservation, but on the ecological utilization and management of lake fishery resources all over the world. The current model was carried out as a case study aimed at (1) modeling the food web structure and energy flows in a typical shallow macrophtyic lake, (2) describing quantitatively the ecosystem properties and maturity of Bao'an Lake, and (3) proposing suggestions for the improvement of fishery resource management in this kind of macrophytic lake.

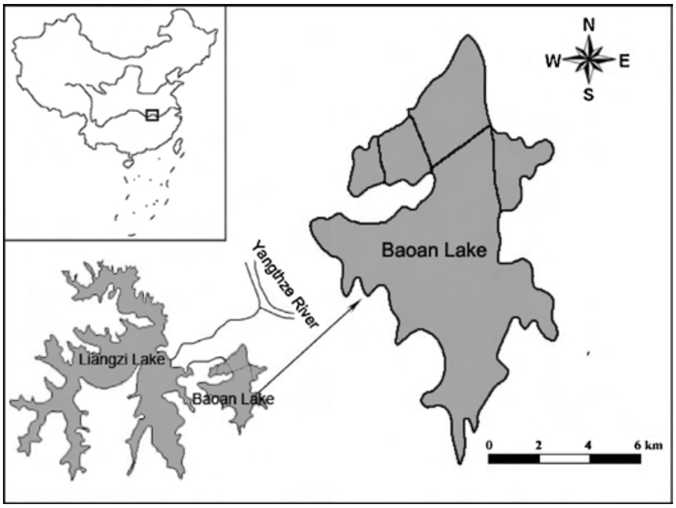


Fig. 1. Geographic location and sketch map of Bao'an Lake.

2. Materials and methods

2.1. Study area

Bao'an Lake (Fig. 1), with a total area of 39.3 km^2, is a typical shallow macrophytic lake (usually with an average water depth of 2.1 m) located on the south bank of the Yangtze River basin within Hubei Province, Central China (Guan, 1995). This lake used to be part of a larger lake, Liangzi Lake, which was connected to the Yangtze River by a channel. According to a survey, the area covered with aquatic macrophytes in the lake reached 75% of the whole lake, dominated by *Vallisneria spiralis* L., *Myriophyllum spicatum* L. and *Nelunbo nucefera* G. (Su et al., 1995). The physicochemical parameters of water quality during the study period are listed in Table 1.

2.2. Ecopath modeling approach

A static mass-balance trophic model was constructed using the Ecopath with Ecosim, version 6.2 (freely available at http://www.ecopath.org; Christensen and Walters, 2004), which focuses on energy transfer between trophic levels and is now widely used in aquatic ecosystem research (Fetahia et al., 2011; Li et al., 2009; Byron et al., 2011; Xu et al., 2011a,b). It differs from other modeling approaches because it encompasses the full trophic spectrum, which makes it appropriate for quantitative assessment of ecosystem structure and function systematically (Christensen, 1995). The basic mass-balance equation of EwE can be described as:

Table 1
Physicochemical parameters of water quality in Bao'an Lake during 1992–1993.

Parameters	Unit	Annual average
Water temperature	°C	18.9
Secchi disk depth	cm	148
pH	–	8.06
Conductivity	μs/cm	267.4
Alkalinity	–	46.54
Dissolved oxygen	mg/L	9.82
Chemical oxygen demand	mg/L	3.60
Total nitrogen	mg/L	1.460
Total phosphorus	mg/L	0.037

Table 2
Species composition of each functional group for the Bao'an Lake ecosystem model.

No.	Functional group	Dominant species composition
1	Mandarin fish	*Siniperca chuatsi*, *Siniperca kneri*
2	Snakehead fish	*Channa argus*
3	Large culters	*Culter alburnus*, *Culter mongolicus*, *Culter dabryi*
4	Catfish	*Pelteobagrus fulvidraco*
5	Common carp	*Cyprinus carpio*
6	Crucian carp	*Carassius auratus*
7	Small fishes	*Pseudorasbora parva*, *Sqalidus nitens*, *Sarcocheilichthys nigripinnis*, *Hemiculter leucisculus*, *Toxabramis swinhonis*, *Gobiidae*
8	Silver carp	*Hypophthalmichthys molitrix*
9	Bighead carp	*Aristichthy nobilis*
10	Grass carp	*Ctenopharyngodon idellus*
11	Breams	*Megalobrama amblycephala*, *Parabramis pekinensis*
12	Crabs	*Eriocheir sinensis*
13	Shrimps	*Caridina nilotic gracilipes*, *Macrobrachium nipponense*, *Neocaridina denticulate sinensis*
14	Molluscs	*Bellamya aeruginosa*, *Alocinma longicornis*, *P. striatulus*
15	Oligochaeta	*Branchiura Sowerbyi*, *Limnodrilus hoffmeisteri*
16	Aquatic insecta	Chironmoidae
17	Microzooplankton	Rorifers, Protozoans
18	Cladocera	*Diaphanosoma leuchtenbergianum*, *Sida crystalline*, *Moina micrura*
19	Copepoda	*Cyclops vicinus*, *Mesocyclops leeckarti*
20	Submerged macrophyte	*Potamogeton maackianus*, *Myriophyllum spicatum*, *Vallisneria spiralis*, *Ceratophyllum demersum*
21	Phytoplankton	Chlorophyta, Bacillariophyta, Cyanophyta
22	Attached algae	Chlorophyta, Bacillariophyta, Cyanophyta
23	Detritus	Bioseston, Abioseston

Production = catches + predation mortality + biomass accumulation + net migration + other mortality or re-expressed concisely and intelligibly as:

$$B_i \cdot \left(\frac{P}{B}\right)_i \cdot EE_i - \sum_{j=1}^{n} B_j \cdot \left(\frac{Q}{B}\right)_j \cdot DC_{ji} - EX_i = 0 \qquad (1)$$

where B_i is the biomass of group i; $(P/B)_i$ represents the production/biomass ratio of group i, which is equal to the coefficient of total mortality Z under steady-state conditions (Allen, 1971); EE_i is the ecotrophic efficiency of group i; B_j is the biomass of predator j; $(Q/B)_j$ is the consumption/biomass ratio of predator j; DC_{ji} is the contribution of prey i in the diet of predator j; EX_i is the export of group i (Christensen and Walters, 2004; Christensen et al., 2005). For each functional group, the composition of the diet and at least three of the four parameters (B, P/B, EE, and Q/B) must be input to establish the model, and the other parameters which are unknown can be estimated by the model (Table 2).

2.3. Data collection and parameter estimation

A total of 23 functional groups were defined to establish the mass-balance model of Bao'an Lake. Groups were classified according to their trophic habits (mainly diet), abundance and the availability of information (Christensen and Walters, 2004). Some commercial fish species and Chinese mitten crab (*Eriocheir sinensis*) were grouped separately due to their importance to fishery yield and stocking (see details in Table 6).

For all functional groups, the key input data used to construct the model usually included biomass (B_i), production/biomass ratio (P_i/B_i), consumption/biomass ratio (Q_i/B_i), ecotrophic efficiency (EE_i) and food composition (DC_{ji}) (Christensen et al., 2005). Normally, fishery capture of some commercial products (such as fish, crabs, shrimps, etc.) was also contained in the model in order to discuss the effects of fishing. Here the fishing yield data were also compiled with the monographs. Because many ecological and fisheries studies had been carried out in Bao'an Lake since the 1980s, especially during the period of 1992–1993, abundant first-hand data on the lake could be collected and compiled from published papers (Hu and Huang, 1991; Liang and Liu, 1995), dissertations (Zhang, 2005; Yan, 1998; Jin, 2003), and unpublished data.

The specific approaches to these parameters are summarized as follows.

2.3.1. Fish

The biomass and P/B of the small fish group in the lake were estimated directly by Zhang (1999) and Zhang (2005). For other fish groups, the biomass and P/B values were calculated from the following equations:

$$B = \frac{Y}{F} \qquad (2)$$

$$F = Z - M \qquad (3)$$

$$Z = \frac{P}{B} = K \cdot \frac{L_\infty - \bar{L}}{\bar{L} - L'} \qquad (4)$$

where B, Y, F, Z and M represent biomass, yield, fishing mortality, total mortality and natural mortality of fish; L_∞, $\bar{L}$ and L' represent the asymptotic length, mean length and cut-off length of fish (Beverton and Holt, 1957; Allen, 1971).

The consumption/biomass ratio (Q/B) of fish was estimated from the empirical equation presented by Palomares and Pauly (1998):

$$\log\left(\frac{Q}{B}\right) = 7.964 - 0.204 \log W_\infty - 1.965T' + 0.083A$$
$$+ 0.532h + 0.398d \qquad (5)$$

where W_∞ is the asymptotic weight (g), T' the mean temperature of the lake expressed by $T' = 1000/\mathrm{K}$ ($\mathrm{K} = {}^\circ\mathrm{C} + 273.15$), A is the aspect ratio ($A = H^2/S$, H is the height of caudal fin and S is the surface area) for a given fish, h is a dummy variable expressing food type (1 for herbivores, and 0 for detritivores and carnivores), and d is a dummy variable also expressing food type (1 for detritivores, and 0 for herbivores and carnivores). The morphometric data for estimating Q/B was derived from Zhang (2005). The diet compositions (DC_{ji}) of fish groups were also modified from Zhang (2005).

2.3.2. Macrocrustaceans

The biomass of the Chinese mitten crab was calculated according to its stocking biomass and recapture rate surveyed during 1991–1993 in the lake (Liu et al., 1995; Jin, 2001), and the biomass of shrimps was estimated directly by Qin et al. (2005) with a pop-net. Since there has been no exact P/B or Q/B data for the macrocrustaceans in the lake until now, we assumed the value of EE and P/Q to be 0.85 and 0.075 respectively according to a similar study in Lake Taihu (Li et al., 2009). The dietary composition came from Jin (2001) for Chinese mitten crab and Li et al. (2009) for shrimps.

2.3.3. Zoobenthos and zooplanktons

An extensive and detailed plankton survey had been carried out during 1992–1993 in Bao'an Lake and the relevant results were compiled in the book "Resource, environment and fishery ecological management of macrophytic lakes" (Liang and Liu, 1995). Therefore, the biomasses of molluscs, oligochaetes, aquatic insecta, microzooplankton, cladocera, copepoda were taken directly from the published materials. The P/B ratio of mollucks, oligochaetes and

Table 3
Basic input and estimated parameters (in bold) for the 23 functional groups of the Bao'an Lake ecosystem model (see details of the abbreviations in Section 2.2 of the text).

Group number	Group name	TL	Biomass ($t\,km^{-2}$)	P/B ($year^{-1}$)	Q/B ($year^{-1}$)	EE	P/Q
1	Mandarin fish	**3.138**	1.44	0.78	3.57	**0.554**	**0.219**
2	Snakehead fish	**3.131**	1.38	1.31	3.46	**0.680**	**0.380**
3	Large culters	**3.143**	4.04	0.96	7.88	**0.797**	**0.122**
4	Catfish	**3.080**	0.32	1.42	7.07	**0.704**	**0.201**
5	Common carp	**2.812**	3.26	1.97	9.17	**0.787**	**0.215**
6	Crucian carp	**2.126**	3.80	2.23	17.02	**0.940**	**0.131**
7	Small fishes	**2.151**	18.00	2.30	8.92	**0.785**	**0.258**
8	Silver carp	**2.021**	2.91	1.15	12.24	**0.970**	**0.094**
9	Bighead carp	**2.315**	3.07	1.02	8.62	**0.858**	**0.118**
10	Grass carp	**2.000**	10.42	1.65	12.41	**0.770**	**0.133**
11	Breams	**2.006**	6.17	1.55	18.31	**0.516**	**0.085**
12	Crabs	**2.227**	2.76	2.12	8.48	**0.770**	**0.250**
13	Shrimps	**2.196**	3.97	4.50	24.40	**0.429**	**0.184**
14	Molluscs	**2.000**	99.10	4.30	**86.00**	**0.126**	0.050
15	Oligochaeta	**2.076**	2.01	5.00	**100.00**	**0.959**	0.050
16	Aquatic insecta	**2.056**	0.72	6.20	**124.00**	**0.913**	0.050
17	Microzooplankton	**2.000**	0.70	**32.02**	**640.37**	0.950	0.050
18	Cladocera	**2.016**	0.23	**48.67**	**973.41**	0.900	0.050
19	Copepoda	**2.016**	0.67	**25.29**	**505.81**	0.900	0.050
20	Submerged macrophyte	**1.000**	1627.00	1.25		**0.120**	
21	Attached algae	**1.000**	41.30	250.70		**0.089**	
22	Phytoplankton	**1.000**	2.43	185.00		**0.376**	
23	Detritus	**1.000**	3.45			**0.645**	

aquatic insecta in the lake were estimated directly by Yan et al. (1998), and the P/Q were assumed to be 0.05 (Yan et al., 2003). For the groups of microzooplankton, cladocera, copepoda, 0.05 and 0.95 were adopted for the values of P/Q and EE (Park et al., 1974; Scavia et al., 1974). All the dietary compositions were modified from Liu (1999) with slight adjustments to balance the model if necessary.

2.3.4. Primary producers and detritus

Phytoplankton, attached algae and submerged macrophytes were defined as the three groups of primary producers. The biomasses of the three groups were investigated directly and reported in the published book (Liang and Liu, 1995). A P/B ratio of 185 $year^{-1}$ for phytoplankton (Sun et al., 1999), 250.7 $year^{-1}$ for attached algae (Wang and Liang, 1995) and 1.25 $year^{-1}$ for macrophytes (Liu, 1992) were used in the same lake. The biomass of detritus was also calculated directly in the lake by Hu and Huang (1991).

2.4. Model balancing and uncertainty

After all the necessary parameters had been entered, the model could be balanced by checking the estimated values: are the EE values possible (less than 1) and are the GE (=P/Q) values physiologically realistic (0.1–0.3) for most groups? Usually, in order to balance a model we modified the input data manually: subtle adjustment should always be made, especially for diet composition (Christensen et al., 2005).

To facilitate this task and to make the process more transparent, we implemented a 'pedigree' routine (Funtowicz and Ravetz, 1990), which serves a dual purpose by describing the origin of the data and by assigning confidence intervals to data based on their origin (Pauly et al., 2000). This index can be understood as a coded statement categorizing the origin of a given input and specifying the likely uncertainty associated with it (Christensen et al., 2005). For each input value, a description should be made on the data source and its confidence (high or low precision, estimated by model, "guesstimated", approximate or indirect method, from other models, or from literature, etc.). Percentage ranges of uncertainty, based on a set of qualitative choices relative to the origin of B, P/B, Q/B, catch and diet composition input were used in the routine and resulted in an index value scaled from 0 (not rooted in local data) to 1 (fully rooted in local data) for each input data point. Based on the individual pedigree index values, an overall 'pedigree index' (P) of the information in ECOPATH was calculated:

$$P = \sum_{i=1}^{n}\sum_{j=1}\frac{l_{ij}}{n} \tag{6}$$

where l_{ij} is the pedigree index for model group i and parameter j, n is the total number of modeled groups (Christensen and Walters, 2004).

Apart from the pedigree index (P), we also used a measure of fit (t^*) to scale the model uncertainty based on the number of living groups in the ecosystem. The measure of fit describes how well a given model is rooted in local data, and is calculated as:

$$t^* = \frac{P * \sqrt{(n-2)}}{\sqrt{(1-P^2)}} \tag{7}$$

3. Results

3.1. Basic input and estimates

After balancing the Ecopath model, a series of estimates and analyses were given by the model. The pedigree index (0.50) lies in the upper range (0.16–0.68) from 150 EwE models (Morissette and Hammill, 2006), and the measure of fit is 2.582 for this model, indicating that the input parameters of the model were based on reliable sources and the model was robust with high confidence (Table 4; Christensen et al., 2000).

The basic input with the estimated parameters given by the EwE for the Baoan Lake ecosystem are shown in Table 3, and the diet matrix for each group is presented in Table 4.

For the Bao'an Lake model, the EE values of all groups were less than 1, and most of the P/Q values were between 0.1 and 0.3, meeting the requirements of a balanced model (Christensen et al., 2005). Generally, the EE values of all the commercial fishery groups were higher than 0.5, and some were even as high as 0.97 (silver carp) and 0.94 (crucial carp). However, the EE of attached algae was extraordinarily low (0.089), followed by submerged macrophytes (0.120) and molluscs (0.126).

Table 4
Diet compositions of 23 functional groups in the Bao'an Lake ecosystem model.

Group	Prey\predator	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	Mandarin fish																			
2	Snakehead fish																			
3	Large culters																			
4	Catfish																			
5	Common carp																			
6	Crucian carp	0.666	0.786																	
7	Small fishes	0.223	0.205	0.790	0.580	0.010							0.150							
8	Silver carp																			
9	Bighead carp																			
10	Grass carp																			
11	Breams																			
12	Crabs																			
13	Shrimps	0.110	0.008	0.193	0.315	0.007														
14	Molluscs		0.001		0.019	0.778		0.132						0.095						
15	Oligochaeta				0.003	0.008	0.007						0.050	0.080						
16	Aquatic insecta			0.001	0.011	0.005	0.008	0.002								0.015				
17	Microzooplankton				0.002				0.006	0.005		0.006		0.005		0.050	0.010		0.016	0.016
18	Cladocera			0.001			0.027	0.008	0.007	0.141				0.003		0.005	0.020			
19	Copepoda						0.081	0.009	0.008	0.164				0.007		0.005	0.025			
20	Submerged Macrophyte										0.977	0.950	0.446							
21	Attached algae						0.050	0.380	0.214	0.200	0.020	0.020	0.149	0.355	0.030	0.380	0.090	0.350	0.537	0.537
22	Phytoplankton						0.022	0.249	0.621	0.383	0.003	0.010	0.005	0.050			0.005	0.150	0.037	0.037
23	Detritus	0.001		0.015	0.070	0.192	0.805	0.220	0.144	0.107		0.014	0.200	0.405	0.970	0.545	0.850	0.500	0.410	0.410
	Sum	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

3.2. Food web structure and trophic analysis

3.2.1. Trophic structure

Fractional trophic levels as suggested by Odum and Heald (1975) were calculated by the Ecopath model. All ecological groups were assigned discrete trophic levels according to Lindeman (1942) with the approach suggested by Ulanowicz (1995). The routine assigns definitional trophic levels (TL) of 1 to producers and detritus and a trophic level of 1 + [the weighted average of preys' TLs] to consumers (Christensen et al., 2005). The fractional trophic levels of all of the 23 groups in Bao'an Lake varied from 1 (primary producer and detritus) to 3.143 (large culters) (Table 3). The primary producers consisted of submerged macrophytes, attached algae and phytoplankton, and the piscivorous fish such as large culter, mandarin fish and snakehead fish occupied the top trophic level in the lake ecosystem.

Two main types of food chains in the Bao'an Lake ecosystem can be discerned from the model: a detrital-based food chain and a grazing food chain. The former chain transferred more energy and matter than the latter one. There were 9055 t km^{-2} year^{-1} matter flowing to trophic level II in the detrital-based food chain, compared with 1333 t km^{-2} year^{-1} in the grazing food chain, although the biomass of the primary producers was extremely high (1671 t km^{-2} year^{-1}) (Fig. 2).

3.2.2. Transfer efficiencies

In order to describe the proportion of energy transferred from one trophic level to the next, all ecological groups in the Bao'an Lake model were assigned to one of eight discrete trophic levels (TL) (Lindeman, 1942). However, only the first five trophic levels were taken into consideration since the flows through TL VI to VIII were extremely small (Fig. 2). The transfer efficiency (TE) of matter is the ratio between the sum of exports and flows predated by the next level and the throughput on the trophic level. For the grazing food chains of the Bao'an Lake ecosystem, the TEs from TL III to TL V were 5.85%, 9.11%, and 12.0% respectively. For the detrital food chain, TEs were 1.11%, 13.3% and 10.8% respectively. A mean TE of 8.96% was calculated for the grazing food chain and 8.40% for the detrital food chain. The geometric mean of the trophic transfer efficiency for the Bao'an Lake ecosystem was 8.68%.

3.2.3. Mixed trophic impacts

A modified input–output analysis with the procedure "Mixed Trophic Impacts (MTI)" described by Ulanowicz and Puccia (1990) was implemented in the EwE. The MTI describes how any group (including fishing fleets) impacts trophically on all the other groups in an ecosystem. It includes both direct and indirect impacts, i.e. both predatory and competitive interactions (Christensen and Walters, 2004). From the mixed trophic impact analysis on the Bao'an Lake ecosystem (Fig. 3), piscivorous fish like mandarin fish and snakehead fish had strong negative effects on crucian carp as well as each other. In addition, large culters also exerted strong negative effects on small fish, shrimps, catfish and themselves, while they had moderate negative effects on mandarin fish and snakehead fish (Fig. 3).

As the main energy sources in the lake ecosystem, detritus and phyplankton had obvious positive effects on many other functional groups. Fishing had strong negative effects on all the commercial fish and crabs, but was beneficial for the small fish and crucian carp (Fig. 3). Nearly all the fish groups had obvious negative effects on forage resources, including forage fish (crucian carps, small fish), zooplankton (microzooplankton, cladocera, copepoda), zoobenthos (molluscs, oligochaeta, aquatic insecta) and algae (attached algae, phytoplankton). On the other hand, the fish groups seemed to have more negative effects on each other mainly due to trophic competition.

3.3. Ecosystem properties and indicators

According to the ecosystem theories implemented by Odum (1969), Odum (1971) and Ulanowicz (1986), a series of indicators that describe ecosystem properties were calculated by the Ecopath to assess the stability and maturity of the ecosystem (Christensen et al., 2005). Summary statistics and flow indices of the Bao'an Lake ecosystem are listed in Table 4.

The total system throughput of the lake ecosystem reached 37,418.040 t km^{-2} year^{-1}, of which 28.2% derived from consumption (10,549.570 t km^{-2} year^{-1}), 13.4% from exports (5009.366 t km^{-2} year^{-1}), 20.9% from respiration (7827.842 t km^{-2} year^{-1}) with 37.5% (14,031.260 t km^{-2} year^{-1}) eventually flowing into detritus. The sum of all production (TP) was 13,449.030 t km^{-2} year^{-1}, and the calculated total net primary

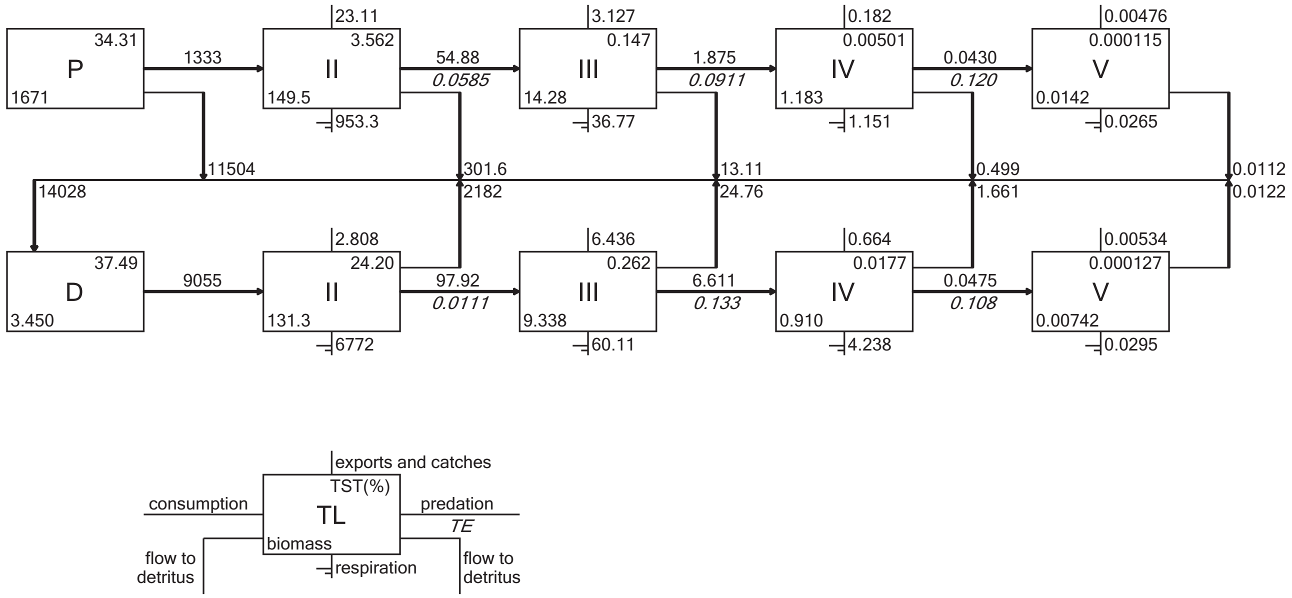


Fig. 2. Lindeman spine of Bao'an Lake ecosystem during 1991–1993 (for details of the units see Table 5).

production (TPP) and the net system production (NSP) were 13,449.030 and 5009.368 t km^{-2} year^{-1} respectively. Thus, the ratio of total primary production/total respiration (TPP/TR) and total primary production/total biomass (TPP/TB) were 1.640 and 6.993 respectively. The mean trophic level of catch was calculated as 2.262, and the gross efficiency (catch/net primary production) was 0.003 in the lake ecosystem.

Flow indices, including connectance index (CI) and system omnivory index (SOI), were used to describe whether the food web is web-like or linear. The values of CI and SOI in the Bao'an Lake ecosystem were 0.205 and 0.058 respectively. Finn's cycling index (FCI; Finn, 1976) and Finn's mean path length (FML) calculated by the model were 9.25% and 2.915 respectively.

Ecosystem information indices, ascendancy (*A*) and system overhead (*O*), are derived from information theory as a measure of the average mutual information in a system (Ulanowicz and Norden, 1990). In the Bao'an Lake ecosystem, the values of ascendancy and overhead were 38.7% and 61.3% respectively (Table 5).

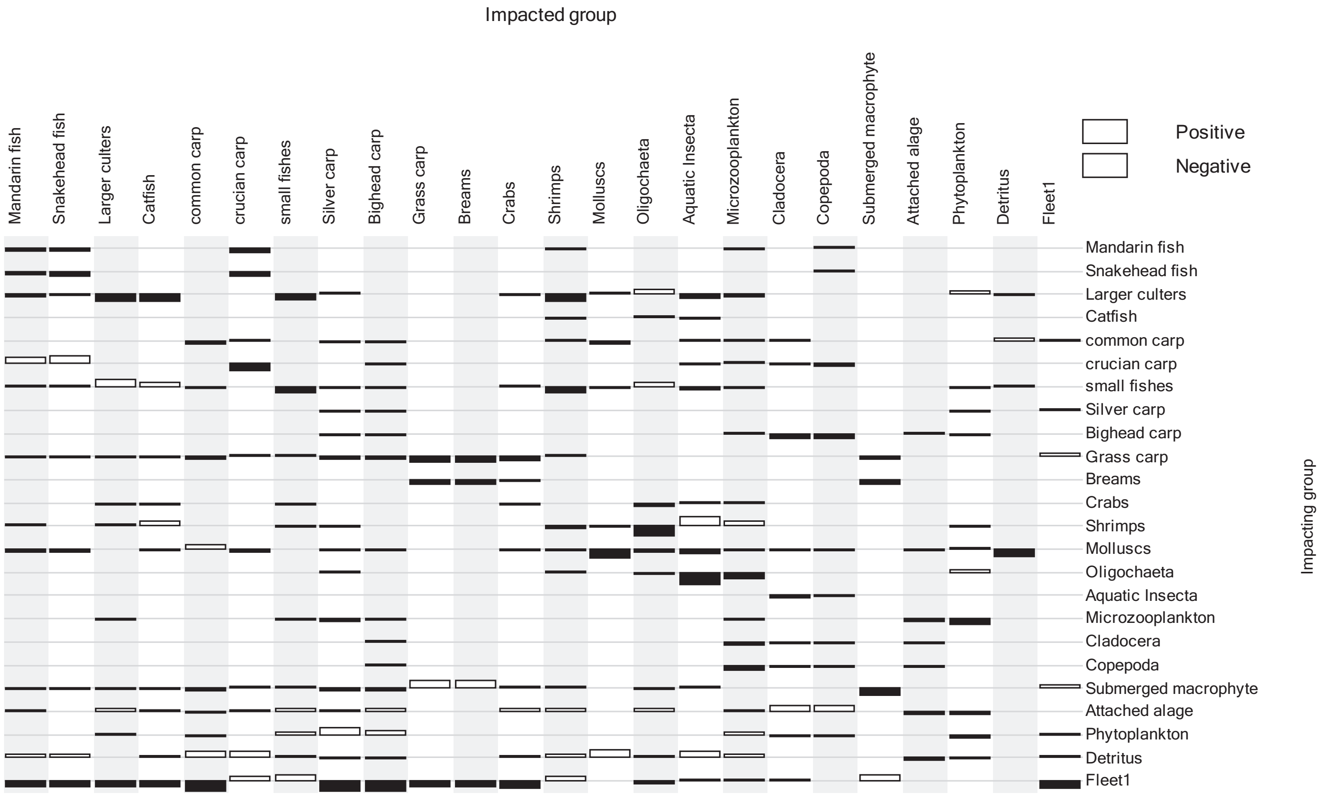


Fig. 3. Mixed trophic impacts of Bao'an Lake ecosystem (white blanks above the line represent a positive impact whereas black blanks underneath the line indicate a negative impact, and the heights of the bars are proportionate to the degree of the impacts).

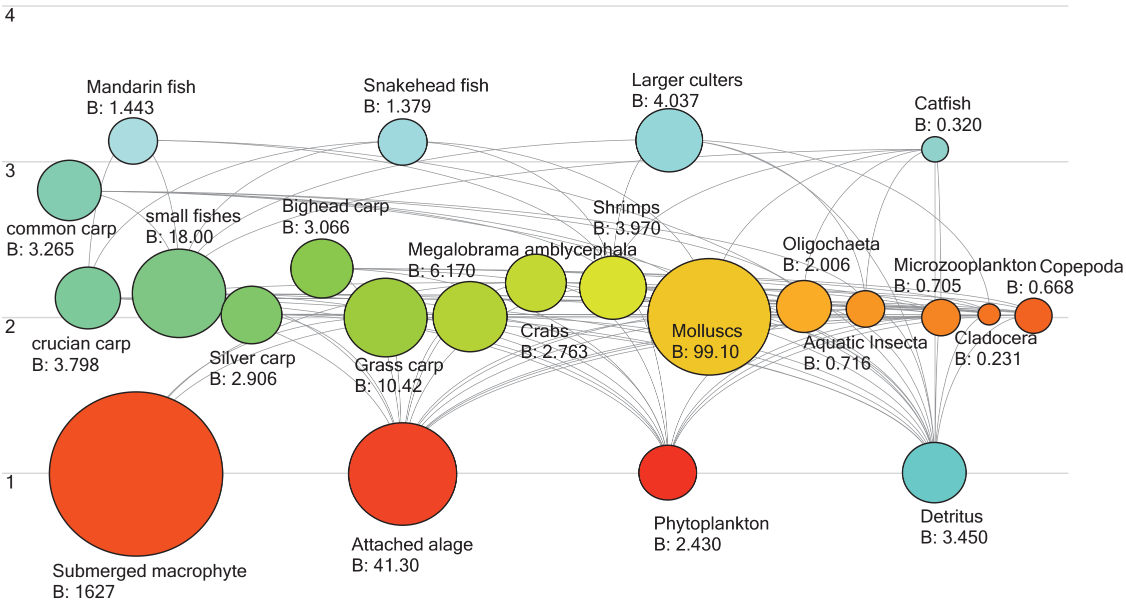


Fig. 4. Schematic diagram of trophic flows and food web structure in the Bao'an Lake ecosystem (for biomass the units are t km^{-2}).

4. Discussion

The current study contributes to the establishment of a mass-balance model to describe the features of food web structure and ecosystem properties in a shallow macrophtyic lake with the objectives of guiding the development of an eco-friendly fishery and the protection and restoration of submerged plants, since the Ecopath with Ecosim model is a systematic and comprehensive model. This model seems to be the first one established in one of the thousands of shallow macrophtyic lakes distributed along the middle reaches of Yangtze River basin. It provides in-depth knowledge of shallow macrophtyic lakes and thus can help to direct us more sustainably in the development and utilization of lake resources.

In the Bao'an Lake ecosystem, the EE values for most fish groups were high, especially for some commercial and stocked fish groups, such as silver carp, bighead carp and grass carp, bream, common carp, mandarin fish and large culters. Additionally, some prey fish groups, such as small fish and crucian carp also showed extremely high EE values. This suggests that commercial fish species are suffering from overexploitation while prey fish suffer from a combination of pressures from predation by piscivores as well as humans, which has lead to a sharp decline in fishery resources (Hu and Huang, 1991). In former times, traditional Chinese carp like silver carp, bighead carp and grass carp were considered to be the main economic stocking species which explains why these fish groups have reached such high biomass in the lake. Another notable feature in the macrophtyic lake is that the biomass of small fish (18 t km^{-2}) and molluscs (99.10 t km^{-2}) is much higher than in other lakes, possibly because submerged plants provide sanctuary for small fish and an attachment matrix for molluscs (Li et al., 2010; Xie et al., 2006), and thus these resources have not been utilized sufficiently due to the traditional fishery model (Fig. 4).

In contrast, the EE values of primary producers, such as submerged plants, attached algae and phytoplankton were as low as 0.12, 0.089 and 0.376 respectively, while the gross efficiency (fishery catch/net primary production) was only 0.003, even lower than Taihu lake (0.0087; Li et al., 2009). This means that primary production was not efficiently utilized and thus very little was transferred into fishery products. Nevertheless, the efficiency of Bao'an Lake is still higher than the global average (0.02%) (Christensen et al., 2005). Large parts of primary production (89.62% of total primary production, about 11,504 t km^{-2} year^{-1}) flowed into detritus. Additionally, the food resource of the Bao'an Lake ecosystem was derived mainly from detritus (87.17%) with little from primary production (12.83%), indicating that the Bao'an Lake ecosystem depended much more on a detritus-based food chain than a grazing food chain, which is also consistent with the suggestion by Odum

Table 5
Summary statistics of the Bao'an Lake ecosystem properties.

Attribute parameters	Value	Units
Sum of all consumption (TC)	10,549.570	t km^{-2} year^{-1}
Sum of all exports (TE)	5009.366	t km^{-2} year^{-1}
Sum of all respiratory flows (TR)	7827.842	t km^{-2} year^{-1}
Sum of all flows into detritus (TD)	14,031.260	t km^{-2} year^{-1}
Total system throughput (TST)	37,418.040	t km^{-2} year^{-1}
Sum of all production (TP)	13,449.030	t km^{-2} year^{-1}
Mean trophic level of the catch (TLc)	2.262	–
Gross efficiency (catch/net primary production)	0.003	–
Calculated total net primary production (TPP)	12,837.210	t km^{-2} year^{-1}
Total primary production/total respiration (TPP/TR)	1.640	–
Net system production (NSP)	5009.368	t km^{-2} year^{-1}
Total primary production/total biomass (TPP/TB)	6.993	–
Total biomass (excluding detritus) (TB)	1835.691	t km^{-2}
Connectance index (CI)	0.205	–
System omnivory index (SOI)	0.058	–
Finn's cycling index (FCI)	9.250	% of total throughput
Finn's mean path length (FML)	2.915	–
Ascendancy (A)	0.387	–
System overhead (O)	0.613	–
Ecopath pedigree index	0.500	–
Measure of fit (t^*)	2.582	–

Table 6
Comparison of ecosystem attributes in different shallow lakes in China.

Parameters	Taihu Lake[a] (1991–1995)	Gehu Lake[b] (1986–1989)	Bao'an Lake[c] (1991–1993)
Mean trophic level of catch (TLc)	2.92	2.78	2.26
Total system throughput (TST)	13,586	12,131.76	37,418.04
Total primary production/total respiration (TPP/TR)	3.85	2.761	1.64
Total primary production/total biomass (TPP/TB)	11.66	1.76	6.99
Ascendancy (*A*)	0.259	0.332	0.387
Connectance index (CI)	0.206	0.208	0.205
System omnivory index (SOI)	0.042	0.086	0.058
Finn's cycling index (FCI)	11.58	14.76	9.25

[a] Li et al. (2009).
[b] Jia et al. (2012).
[c] Present study.

(1969) that a mature system may depend more on the detrital pathway. The fact that the Bao'an Lake food web relies on two energy sources may enhance the plasticity and resilience of the ecosystem (Fetahia et al., 2011). According to the MTI analysis, the two main food sources in the Bao'an Lake ecosystem, i.e. detritus and primary producers (especially attached algae and phytoplankton), had significant positive effects on other groups. Other important information from the MTI is that fishing pressure may have more negative impacts on commercial fish groups but will be better for small fish resources, consistent with the fact that fish tend to miniaturization (Cao et al., 1991). The MTI also showed that fishing pressure exerted stronger impacts on most functional groups than predation or competition (Christensen and Walters, 2004).

The mean transfer efficiency among different trophic levels of the ecosystem was 8.68% in our study, slightly lower than the 10% assumed by Lindeman (1942), but lying in the accepted range of EE values reported in the published literature (Libralato et al., 2008; Pauly and Christensen, 1995). The low transfer efficiency mainly results from transfer from TL II to TL III in the food chain. From the model, it is evident that a large amount of mollusc biomass was not utilized efficiently.

As previously mentioned, no such model has been established for the lakes along the middle reaches of the Yangtze River, although there are some preliminary studies in some large lakes nearby: Taihu lake (Li et al., 2009) and Gehu lake (Jia et al., 2012) (Table 6). A comparison with these three lake ecosystems on some key parameters from network analysis would be helpful for accurately positioning and characterizing the Bao'an Lake ecosystem (Table 6). Notably, the total system throughput in Bao'an Lake (37,418.04 t km^{-2} $year^{-1}$) is much higher than in Taihu Lake (13,586 t km^{-2} $year^{-1}$) or Gehu lake (12,131.76 t km^{-2} $year^{-1}$). This is mainly because the biomass of primary producers, such as submerged plants accompanied by attached algae was much higher than in the other two lakes, indicating that Bao'an Lake is typically a lake dominated by submerged plants.

According to Odum (1971), the ratio of total primary production to total respiration (TPP/TR) is an important measure of ecosystem maturity; ecosystems with the values much higher or lower than 1 are thought to be immature, while only those with TPP/TR ratios approaching 1 are considered to be mature. In our study, the value of the TPP/TR ratio was 1.640, much lower than the 3.85 in Taihu or 2.761 in Gehu (Li et al., 2009; Jia et al., 2012), suggesting that the Bao'an Lake ecosystem was more mature compared with the other two lakes. Additionally, as modified from Odum (1969), the lower TPP/TB ratio also characterizes a mature ecosystem. The ratio in this study was 6.99, between the immature ecosystem (Taihu Lake, 11.66) and the mature ecosystem (Gehu Lake, 1.76). In Bao'an Lake, ascendancy (38.7%), a measure of ecosystem growth and development, is much higher than in Taihu Lake (25.9%) or Gehu lake (33.2%) which also suggests a mature system.

FCI represents the fraction of an ecosystem's throughput that is recycled compared to total system throughput (Finn, 1976; Table 3). The value of 9.25% for Bao'an Lake is much lower than in the other two lakes (Taihu Lake, 11.58%; Gehu Lake, 14.76%). Meanwhile, CI and SOI are two other important indices which describe system maturity since the food chain is expected to change from linear to web-like as the system matures (Odum, 1971). However, in Bao'an Lake, we can see that the value of CI and SOI (0.205 and 0.058 respectively) were both relatively low. Although these values are nearly in the same order of magnitude as some mature systems such as Gehu (Jia et al., 2012), Qiandaohu (Liu et al., 2007) and Taihu lakes (Li et al., 2009), there is still a large gap (especially SOI) with some other mature systems such as Hayq Lake (SOI = 0.224; Fetahia et al., 2011), Lake Toya (SOI = 0.12; Hossain et al., 2010) and Lake Kuvi (SOI = 0.148; Villanueva et al., 2008).

Consequently, the high values of TPP/TR and ascendancy and the low values of TPP/TB, and the more important detritus dominated food chain, all illustrate that the Bao'an Lake ecosystem is a mature and stable system according to the theories of Odum (1969). However, the moderate value of FCI along with extremely low values of CI and SOI still shows that the food web structure of the lake tends to lack complexity (Odum, 1969). Thus we conclude that the Bao'an Lake ecosystem is a mature system but with a simple and vulnerable food web structure. This is mainly caused by unsustainable fish stocking and overfishing, since we know that prior to our study, piscivorous fish were targeted for removal from the lake, so that more traditional carp could be released. This resulted in significant changes in the fish community structure (Zhang et al., 1997, unpublished data). It has resulted in the mean trophic level of catch in Bao'an Lake being only 2.26, far lower than Taihu Lake (2.92) or Gehu Lake (2.78), which indicates that Bao'an Lake is suffering from overfishing and unsustainable fish stocking with too many low trophic level fish fry being released and caught in the lake leading to the simple food web structure.

5. Conclusions and suggestions for fishery management improvement in macrophyte-dominated lakes

Generally, this is the first ecosystem model of a shallow macrophytic lake in the middle reaches of the Yangtze River basin, which can quantitatively describe the food web structure and ecosystem properties. Based on the ecosystem analysis in Bao'an Lake, we suggest that fish stocking in this lake should aim at increasing species richness, because numerous studies have showed that predator diversity can strengthen ecosystem function and food web structure (Griffin et al., 2008; Carey and Wahl, 2011; Hargrave, 2009). We suggest that the lake be stocked with piscivorous and omnivorous species since these two groups seemed to be the key factors in mediating biodiversity–ecosystem functioning relationships (Petchey et al., 2004; Bruno and O'Connor, 2005). This fishery practice can also fully utilize the resources in this type of lake and increase the system's SOI. This suggestion is also consistent with the former studies in these kinds of lakes (Xie et al., 2000; Cui and Li, 2005), which recommended that mandarin fish (*S. chuatsi*) and Chinese mitten crab (*E. sinensis*) can be reasonably stocked.

Additional attention should be given to the utilization of macrophytes. Some studies have indicated that dense vegetation could result in low feeding efficiency and poor growth of both small fish and piscivorous fishes (Colle and Shireman, 1980; Miranda and Pugh, 1997; Xie et al., 2005). Thus, herbivorous species with high economic value such as Chinese mitten crab (*E. sinensis*) and *Megalobrama amblycephala* Yih could be stocked moderately (Cui and Li, 2005).

It should be acknowledged that the Ecopath model is just a steady-state model and therefore it cannot forecast the dynamic effects of fish stocking on the other organisms within the food web. The current study is just part of ongoing studies in these lakes, and further studies should be implemented to probe the carrying capacity of multi-species stocking based on the whole food web and ecosystem using the dynamic model (Ecosim). That will give more credible direction and prediction of the effects of fish stocking in this kind of lake.

Acknowledgements

We greatly appreciated Prof. Jane Hughes and Miss Chintha Perera for their kind help and suggestions in manuscript correcting. Thanks to the editor and the two anonymous referees for their valuable comments and helpful corrections. This research was financially supported by Grant Nos. 200903048 and 20130305 from the "Special Fund for Agro-scientific Research in the Public Interest" of China, Nos. 30830025 and 30900182 from the National Natural Science Foundation of China, No. 2012BAD25B08 from the National Science and Technology Supporting Program of China and No. 2011FBZ28 from the State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences. We are grateful to China Scholarship Council (CSC) for their financial support of study abroad.

References

Allen, K.F., 1971. Relation between production and biomass. Journal of the Fisheries Research Board of Canada 20, 1573–1581.

Beverton, R.J.H., Holt, S.J., 1957. On the Dynamics of Exploited Fish Populations. Fishery Investment, London.

Bruno, J.F., O'Connor, M.I., 2005. Cascading effects of predator diversity and omnivory in a marine food web. Ecology Letters 8, 1048–1056.

Byron, C., Link, J., Costa-Piercea, B., Bengtson, D., 2011. Calculating ecological carrying capacity of shellfish aquaculture using mass-balance modeling: Narragansett Bay, Rhode Island. Ecological Modelling 222, 1743–1755.

Carey, M.P., Wahl, D.H., 2011. Fish diversity as a determinant of ecosystem properties across multiple trophic levels. Oikos 120, 84–94.

Cao, W.X., Zhang, G.H., Ma, J., Yu, D.H., 1991. Preliminary studies on the phenomenon of size diminution of the fish resources in Lake Honghu. In: Honghu Research Group, Institute of Hydrobiology, CAS (Eds.), Studies on Comprehensive Exploitation of Aquatic Biological Productivity and Improvement of Ecological Environment in Lake Honghu. China Ocean Press, Beijing, pp. 148–152 (in Chinese with English abstract).

Chen, H., 1989. Impact of aquaculture on the ecosystem of the Donghu Lake, Wuhan. Acta Hydrobiologica Sinica 13 (4), 359–368 (in Chinese with English abstract).

Christensen, V., 1995. Ecosystem maturity—towards quantification. Ecological Modelling 77, 3–32.

Christensen, V., Walters, C.J., 2004. Ecopath with Ecosim: methods, capabilities, and limitation. Ecological Modelling 172, 109–139.

Christensen, V., Walters, C.J., Pauly, D., 2005. Ecopath with Ecosim: a user's guide. Fisheries Centre, University of British Columbia, Vancouver, pp. 154.

Colle, D.E., Shireman, J.V., 1980. Coefficients of condition for largemouth bass, bluegill, and redear sunfish in hydrilla-infested lakes. Transactions of the American Fisheries Society 109 (5), 521–531.

Cui, Y.B., Li, Z.J., 2005. Fishery Resources and Conservation of Environment in Lakes of the Changjiang River Basin. Science Press, Beijing.

Fetahia, T., Schagerl, M., Mengistoua, S., Libralato, S., 2011. Food web structure and trophic interactions of the tropical highland lake Hayq, Ethiopia. Ecological Modelling 222, 804–813.

Finn, J.T., 1976. Measures of ecosystem structure and function derived from analysis of flows. Journal of Theoretical Biology 56, 363–380.

Funtowicz, S.O., Ravetz, J.R., 1990. Uncertainty and Quality in Science for Policy. Kluwer Academic Publishers, Dortrecht.

Griffin, J.N., Haye, K.L., Hawkins, S.J., Thompson, R.C., Jenkins, S.R., 2008. Predator diversity and ecosystem function: density modifies the effect of resources partition. Ecology 89 (2), 298–305.

Guan, Z.H., 1995. Morphometric parameters of Bao'an lake and their limnological implications. In: Resources, Environment and Fishery Ecological Management of Macrophytic Lakes. Science Press, Beijing, pp. 3–15 (in Chinese with English abstract).

Heymans, J.J., Lynne, J., Shannon, Astrid, J., 2004. Changes in the northern Benguela ecosystem over three decades: 1970s, 1980s, and 1990s. Ecological modelling 172 (2), 175–195.

Hu, C.L., Huang, X.F., 1991. Collected Papers on the Fishery Ecology and Exploitation Technology of the Bao'an Lake. Science Press, Beijing (in Chinese with English abstract).

Hossain, Md.M., Matsuishi, T., Arhonditsis, G., 2010. Elucidation of ecosystem attributes of an oligotrophic lake in Hokkaido, Japan, using Ecopath with Ecosim (EwE). Ecological Modelling 221, 1717–1730.

Hargrave, C.W., 2009. Effects of fish species richness and assemblage composition on stream ecosystem function. Ecology of Freshwater Fish 18, 24–32.

Jia, P.Q., Hu, M.H., Hu, Z.J., Liu, Q.G., Wu, Z., 2012. Modeling trophic structure and energy flows in a typical macrophyte dominated shallow lake using the mass balanced model. Ecological Modelling 233, 26–30.

Jin, G., 2001. Studies on the biology and fishery management of Chinese Mitten Crab, Ericheir sinensis. Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan (PhD dissertation) (in Chinese with English abstract).

Li, W., Zhang, T.L., Li, Z.J., 2010. Spatial distribution and abundance of small fishes in Xiaosihai Lake, a shallow lake along the Changjiang (Yangtze) River, China. Chinese Journal of Oceanology and Limnology 28 (3), 470–477.

Li, Y.K., Chen, Y., Song, B., Olson, D., Yu, N., Chen, L.Q., 2009. Ecosystem structure and functioning of Lake Taihu (China) and the impacts of fishing. Fisheries Research 95, 309–324.

Liang, Y.L., Liu, H.Q., 1995. Resources Environment and Fishery Ecological Management of Macrophytic Lakes. The Science Press, Beijing (in Chinese with English abstract).

Libralato, S., Coll, M., Tudela, S., Palomera, I., Pranovi, F., 2008. Novel index for quantification of ecosystem effects of fishing as removal of secondary production. Marine Ecology Progress Series 355, 107–129.

Liu, J.K., 1992. Freshwater Aquaculture. Technology press, Beijing.

Liu, H.Q., Li, Z.J., Fang, R.L., 1995. A series of studies on high quality and high efficiency ecological fishery model of Bao'an Lake. In: Resources, Environment and Fishery Ecological Management of Macrophytic Lakes. The Science Press, Beijing, pp. 228–236 (in Chinese with English abstract).

Liu, J.K., 1984. Lakes of the middle and lower basins of the Changjiang (China). In: Taub, F.B. (Ed.), Ecosystems of the World 23 Lakes and Reservoirs. Elsevier, Amsterdam, pp. 331–355.

Liu, Q.G., Chen, Y., Li, J.L., Chen, L.Q., 2007. The food web structure and ecosystem properties of a filter-feeding carps dominated deep reservoir ecosystem. Ecological Modelling 203, 279–289.

Liu, J., He, B., 1992. Cultivation of the Chinese Freshwater Fishes. The Science Press, Beijing, pp. 381 (in Chinese with English abstract).

Liu, J.K., 1999. Advanced Aquatic Biology. Science Press, Beijing.

Lindeman, R.L., 1942. The trophic-dynamic aspect of ecology. Ecology 23, 399–417.

Miranda, L.E., Pugh, L.L., 1997. Relationship between vegetation coverage and abundance, size, and diet of juvenile largemouth bass during winter. North American Journal of Fisheries Management 17 (3), 601–610.

Morissette, L., Hammill, M.O., Savenkoff, C., 2006. The trophic role of marine mammals in the northern Gulf of St Lawrence. Marine Mammal Science 22, 74–103.

Odum, E.P., 1969. The strategy of ecosystem development. Science 164, 262–270.

Odum, E.P., 1971. Fundamental of Ecology. Saunders, Philadelphia.

Odum, W.E., Heald, E.J., 1975. The detritus-based food web of an estuarine mangrove community. In: Cronin, L.E. (Ed.), Estuarine Research. Academic Press, New York, pp. 265–286.

Petchey, O.L., Downing, A.L., Mittelbach, G.G., Persson, L., Steiner, C.F., Warren, P.H., et al., 2004. Species loss and the structure and functioning of multitrophic aquatic systems. Oikos 104, 467–478.

Park, R.A., O'Neill, R.V., Bloomfield, J.A., 1974. A generalized model for simulating lake ecosystems. Simulation 23, 33–50.

Pauly, D., Christensen, V., 1995. Primary production required to sustain global fisheries. Nature 374, 255–257.

Pauly, D., Christensen, V., Walters, C., 2000. Ecopath, Ecosim, and Ecospace as tools for evaluating ecosystem impact of fisheries. Journal of Marine Science 57, 697–706.

Palomares, M.L.D., Pauly, D., 1998. Predicting food consumption of fish populations as functions of mortality, food type, morphometrics, temperature and salinity. Marine and Freshwater Research 49 (5), 447–453.

Qin, H.M., Zhang, T.L., Li, Z.J., Hong, Y.J., 2005. Species composition, spatial distribution and biomass of shrimp community in the Biandantang Lake. Acta Hydrobilogica Sinica 29 (4), 379–384.

Scavia, D., Bloomfield, J.A., Fisher, J.S., 1974. Documentation of CLEANX: a generalized model for simulating the open-water ecosystems of lakes. Simulation 23, 51–56.

Su, Z.G., Zhang, T.L., Cai, Q.H., 1995. On change of aquatic vegetation in Bao'an Lake with remarks on its fishery utilization. In: Resources, Environment and Fishery Ecological Management of Macrophytic Lakes. The Science Press, Beijing, pp. 147–159 (in Chinese).

Tong, L., 1999. Ecopath model—a mass-balance modeling for ecosystem estimation. Marine Fisheries Research 20 (2), 103–107 (in Chinese with English abstract).

Ulanowicz, R.E., 1995. Ecosystem trophic foundations: Lindeman exonerata. In: Patten, B.C., Jørgensen, S.E. (Eds.), Complex Ecology: The Part–Whole Relation in Ecosystems. Prentice Hall, Englewood Cliffs, pp. 549–560 (Chapter 21).

Ulanowicz, R.E., 1986. Growth and Development: Ecosystem Phenomenology. Springer Verlag, New York.

Ulanowicz, R.E., Puccia, C.J., 1990. Mixed trophic impacts in ecosystem. Coenoses 5, 7–16.

Villanueva, M.C., Isumbisho, M., Kaningini, B., Moreau, J., Micha, J.C., 2008. Modeling trophic interactions in Lake Kivu: what roles do exotics play? Ecological Modelling 212 (4), 2–438.

Wang, J., Liang, Y.L., 1995. A preliminary study on the structure and abundance of the attached algae in Bao'an lake. In: Resources, Environment and Fishery Ecological Management of Macrophytic Lakes. The Science Press, Beijing (in Chinese), pp. 228–236.

Xie, S.G., Cui, Y.B., Zhang, T.L., Fang, R.L., Li, Z.J., 2000. The spatial pattern of the small fish community in the Biandantang Lake—a small shallow lake along the middle reach of the Yangtze River, China. Environmental Biology of Fishes 57, 179–190.

Xie, P., Chen, Y., 1996. 'Evil Quarter' of inland waters in China—impact of human activities on the loss of biodiversity. Acta Hydrobiologica Sinica 20 (Suppl.), 6–23 (in Chinese with English abstract).

Xu, S.N., Chen, Z., Li, S., He, P., 2011a. Modeling trophic structure and energy flows in a coastal artificial ecosystem using mass-balance Ecopath model. Estuaries and Coasts 34, 351–363.

Xu, S.N., Chen, Z., Li, C., Huang, X.P., Li, S.Y., 2011b. Assessing the carrying capacity of tilapia in an intertidal mangrove-based polyculture system of Pearl River Delta, China. Ecological Modelling 222, 846–856.

Zhang, T.L., 2005. Life-history strategies, trophic patterns and community structure in the fishes of Lake Biandantang. Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan (PhD thesis) (in Chinese with English abstract).

Large-scale patterns of fish species distribution and assemblage: a novel multi-species modeling approach

Chuanbo Guo[1,2,3], Sovan Lek[1,2], Shaowen Ye[1*], Wei Li[1], Jiashou Liu[1], Tanglin Zhang[1], Jing Yuan[1] and Zhongjie Li[1*]

[1] State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China

[2] Université de Toulouse, UMR 5174 EDB, CNRS-Université de Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex 09, France

[3]University of Chinese Academy of Sciences, Beijing 100039, China

*Correspondance: Shaowen Ye & Zhongjie Li, State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China.

Email: yeshw@ihb.ac.cn; zhongjie@ihb.ac.cn

Running head: Large-scale patterns of fish diversity and distribution

[Submitted] *Freshwater Biology*

Summary

1.This work contributed to the building of the first global perspective on the fish distribution and assemblage patterns, as well as the understanding of the potential drivers in a large geographical scale across China, using a novel multi-species approach fitted by the multivariate regression tree.

2.A total of 425 fish species and subspecies recorded in 135 lakes were modeled with 19 potential determinants. Finally, five fish assemblages were defined by the constrained clustering, 107 indicator species were thus identified. Species diversity showed significantly differences among each assemblage: fish species richness in plateau lakes was significantly lower than plain lakes; however the diversity of the whole assemblage in plateaus was higher than other regions.

3.Altitude, minimum temperature of the coldest month, annual temperature range and precipitation during the driest month were found to be the most important determinants affecting fish assemblages and distribution patterns in Chinese lakes.

4.Overall, our results highlighted the plausibility of the species-energy hypothesis rather than species-area hypothesis for fish species distribution patterns in large scale. In addition, we emphasized that climate change will severely affect the lake fish diversity and distribution. These findings could have important implications for fish species conservation and natural resources management all over the world.

Keywords: China, Distribution patterns, Fish assemblage, Geographical and climatic variables, Multivariate regression tree

Introduction

During the last several decades, ongoing global climate change, environment pollution and anthropogenic disturbance have contributed to a dramatic change in biodiversity all over the world (DeFries *et al.,* 2004; Fang *et al.,* 2009). This situation is very significant for fish, and likely to continue in the future, as for them adaptation to the pressure from complex global changes has proven difficult (Buisson *et al.,* 2008; Magurran, 2009). Therefore, identifying the determinants that affect the patterns of fish assemblages and distributions becomes a global research hotspot in ecology, biogeography, evolution and conservation sciences (He *et al.,* 2011; Park *et al.,* 2006; Sharma *et al.,* 2007; Buisson *et al.,* 2008; Grenouillet *et al.,* 2011; Maloney *et al.,* 2013). Furthermore, detailed knowledge of species' ecological and geographic distributions is fundamental for effective fish conservation and forecasting (Rushton *et al.,* 2004). However, for a long time, it is too difficult to assess and monitor organisms in a large scale, while in this case, species distribution models (SDM) which could integrate fragmental information of species and habitats globally showed it's advantages (Guisan and Zimmermann, 2000). Species distribution models (SDMs) are generally developed to quantify the association between species' occurrence or abundance and environmental, climatic or geographical predictors (*e.g.* Guisan & Zimmerman, 2000; Araújo & Guisan, 2006; Elith *et al.,* 2007), and different kinds of SDMs have been widely used for (1) predicting distribution patterns or assemblages from the current habitat status (*e.g.* Park *et al.,* 2006; Buisson *et al.,* 2008; Grenouillet *et al.,* 2011), (2) evaluating the potential spreading capacity of invasive species (*e.g.* Roura-Pascual *et al.,* 2009; Brummer *et al.,* 2013), and (3) assessing biological responses and occurrences to global changes (*e.g.* Thuiller *et al.,* 2005; Araujo *et al.,* 2006; Buisson & Grenouillet, 2009; Tisseuil *et al.,* 2012; Yu *et al.,* 2013). With the rapid development of computing capacity and large databases, the SDMs have been widely used in both basic and applied ecology and increasing novel and robust models have emerged. For instance, multivariate regression trees (MRT), one form of a multi-response regression tree model (De'ath, 2002), can be regarded as a constrained clustering method with each cluster described by a set of environmental variables and representing an assemblage type. The most important advantage of the MRT analysis over the individual species analyses, such as generalized linear models (GLM), generalized additive models (GAM), classification and regression trees (CART), and random forest (RF), is the capacity and facility in modeling and predicting multi-species data. Simultaneously, regression tree analysis has long proved to give more accurate and robust predictions than linear models, especially for analyzing complex ecological data (De'ath & Fabricius 2000; Vayssieres *et al.*, 2000). To date, MRT has been widely used as a constrained cluster method (DeVantier *et al.*, 2006; Davidson *et al.*, 2010; Hamann *et al.*, 2011).

In aquatic ecosystems, a growing concern has highlighted the application of different SDMs or even ensemble models to describe and predict the patterns of fish assemblages and distributions with diverse determinants (e.g. Park *et al.*, 2006; Buisson *et al.*, 2008; He *et al.*, 2011; Maloney *et al.*, 2013).However, most of these studies only focused on a small geographical range, in some rivers or streams. Large-scale patterns of fish diversity and distribution still lack of global insights and studies. Moreover, compare with other aquatic ecosystems, lakes remain poorly studied despite the fact that they are among the most important aquatic ecosystems on earth and can be ideal fields for studying fish distribution and assemblage since they are natural bio-geographical islands (Wang & Dou, 1998; Zhao *et al.*, 2006). Chinese lakes, occupying significantly different natural environments, climatic factors and fish biota due to the vast area, have always been considered as the 'centre of dispersal' for fish species (Wu, 1964; 1977). More than 400 fish species and subspecies live in these lakes (Wang & Dou, 1998), making them more than suitable for the case studies of large-scale patterns of fish diversity and distributions, as well as the relationships with environment and climatic variations. However, for a long time, fish studies in these lakes have been focused on population biology (*e.g.* Zhang *et al.*, 2005), community structure (*e.g.* Xie *et al.*, 2001; Ye *et al.*, 2006) and fishery management (*e.g.* Cui & Li, 2005) in single or several lakes on regional scale, while a few studies focused on the national scale but with limited data and analysis (Zhao *et al.*, 2006; Kang *et al.*, 2013).

Up until now, there is nearly no global understanding of fish assemblages and distribution patterns with the potential determinants in Chinese, especially in large scale. Conversely, fish fauna in the lakes across China have been extensively surveyed from the 1950s, and the improvement in statistical and modeling approaches are routinely available nowadays. Thus a more comprehensive and complete picture of the fish assemblages and distribution patterns as well as their driving factors with a global perspective is urgently required, which in turn serves to benefit biodiversity conservation and management and the health of ecosystems.

Therefore, in this context, a multi-response model fitted by a multivariable regression tree on a national scale was constructed, using more reliable and ideal fish presence-absence data (Cawsey *et al.,* 2002; Elith *et al.,* 2007). Generally, our research contributed to draw up the first global perspective on fish diversity and distribution patterns in China, identify the importance of factors in driving the fish diversity, verify the hypothesis appropriated for fish distribution and assemblage patterns, and provide eco-solutions for the conservation and management of fish biodiversity and natural resources.

Methods

Study area

Lakes in China always had with complicated natural environments and large geological variations. In the present study, a total of 135 lakes (21.9-48.9º N and 81.2-132.6º E) distributed across China with creditable fish presence-absence data were selected. The surface areas of these lakes ranged from 0.0006 to 2933 km^2. The detailed location and limnological characteristics of each lake are summarized in Table 1.

Fish data

The presence-absence data of fish in the 135 lakes were compiled from published fish surveys since the 1950s (see Appendix S1 for details). To ensure the quality and authenticity of the data, fish compositions derived from long-term monitoring and surveying were adopted, while the data with insufficient sampling were omitted. For the lakes that had fish fauna data from several time periods, normally the earliest were preferred as they may have suffered from fewer artificial disturbances. Fish composition data were then revised according to the monograph “*Fauna Sinica*” and Wu (1977; 1964) to deal with possible nomenclatural changes (e.g. synonyms, newly discovered and named fish species). Consequently, 425 fish species and subspecies were included.

Climatic and environmental variables

A set of climatic and lake geographic variables were used as the explanatory variables. The climatic variables were derived from the WordClim database (Hijimans *et al.*, 2005, available at http://www.worldclim.org) using Arcgis 10.1 (ESRI). This database was developed from compiled monthly averages of climate as measured at weather stations from a large number of global, regional, national and local sources, mostly from the 1950–2000 period, using the Thin Plate

Table.1 Comprehensive description of all the variables used in this study

Variables	Variable type	Min	1st Qu	Median	Mean	3rd Qu	Max	SD
Latitude	Mean latitude	21.90	28.45	31.10	34.42	41.67	48.90	7.90
Longitude	Mean Longitude	81.20	103.00	114.40	111.60	118.40	132.60	10.47
alt	Elevation of the site	2.00	23.50	144.00	911.70	1773.00	4947.00	1154.56
area	Surface area of the lake	0.01	6.72	27.95	225.80	122.33	4930.00	544.82
bio1	Annual mean temperature	-5.00	50.00	140.00	113.70	170.00	217.00	61.45
bio2	Mean diurnal range: mean of monthly (max temp–min temp)	70.00	82.00	110.00	104.50	118.00	146.00	19.95
bio3	Isothermality: (Bio2/Bio7) *100	21.00	24.00	25.00	30.16	36.50	53.00	9.78
bio4	Temperature seasonality (SD *100)	3242.00	6903.00	8807.00	9305.00	12168.00	15770.00	3468.65
bio5	Maximum temperature of warmest month	130.00	258.00	286.00	284.40	323.50	340.00	46.71
bio6	Minimum temperature of coldest month	-295.00	-196.50	-22.00	-79.59	9.50	93.00	109.16
bio7	Temperature annual range (Bio5–Bio6)	212.00	306.50	331.00	364.00	452.50	561.00	103.52
bio8	Mean temperature of wettest quarter	78.00	199.50	216.00	214.30	244.50	283.00	40.57
bio9	Mean temperature of driest quarter	-208.00	-117.00	50.00	-2.47	79.50	162.00	104.59
bio10	Mean temperature of warmest quarter	78.00	202.00	221.00	227.00	272.00	286.00	48.12

bio11	Mean temperature of coldest quarter	-213.00	-119.50	43.00	-12.32	57.00	149.00	99.40
bio12	Annual precipitation	82.00	418.50	919.00	806.30	1177.50	1480.00	409.61
bio13	Precipitation of wettest month	18.00	132.50	186.00	168.60	225.00	305.00	63.03
bio14	Precipitation of driest month	0.00	2.00	9.00	15.04	31.50	44.00	14.62
bio15	Precipitation seasonality(coefficient of variation)	39.00	55.00	86.00	83.35	106.50	138.00	26.64
bio16	Precipitation of wettest quarter	47.00	291.00	461.00	423.70	573.00	815.00	171.80
bio17	Precipitation of driest quarter	2.00	8.00	38.00	56.64	122.00	160.00	55.12
bio18	Precipitation of warmest quarter	47.00	291.00	443.00	403.20	516.00	676.00	157.02
bio19	Precipitation of coldest quarter	2.00	8.00	39.00	60.01	129.50	175.00	59.59

Smoothing Spline (TPS) algorithm that yielded climate surfaces for monthly maximum, minimum, mean temperatures and total monthly precipitation (Hutchinson, 1995). In order to improve the accuracy of our analyses, we choose the highest resolution (30 arc-seconds (~1 km)). A PCA (principal component analysis) was performed to eliminate the variables with high correlation, and finally 19 variables are included.

Data analysis

MRT model

In our study, a novel technique for modeling species-environment relationships, MRT, was used as the main tree model to explore the relationships between fish species and environments. MRT is the tree model based on the same recursive partitioning principles as Classification and Regression Trees (CART) but extended to multivariate response variables, which is an important consideration for the community studies. MRT splits objects (e.g. sampling sites) into homogenous groups according to the response, with the splits constrained by explanatory variables. The tree is grown by splitting the data a large number of times, and then it is subsequently pruned (reduction of the number of groups) via a re-sampling method called v-fold cross-validation (Breiman *et al.*, 1984) to obtain the best predictive tree size. The model was run with 100 replicates to be certain that the results were not simply obtained by chance. An unconstrained cluster (hclust) was also presented to compare with the MRT groups, a Kruskall-Wallis test and multiple comparison tests were then conducted in order to assess the differences in species richness between each assemblage (Giraudoux, 2006). Here, the new R package “MVPART wrap” (Ouellette *et al.*, 2012) was used instead of the package “mvpart” (De’ath, 2002) to get more detailed information from the MRT model.

Indicator species

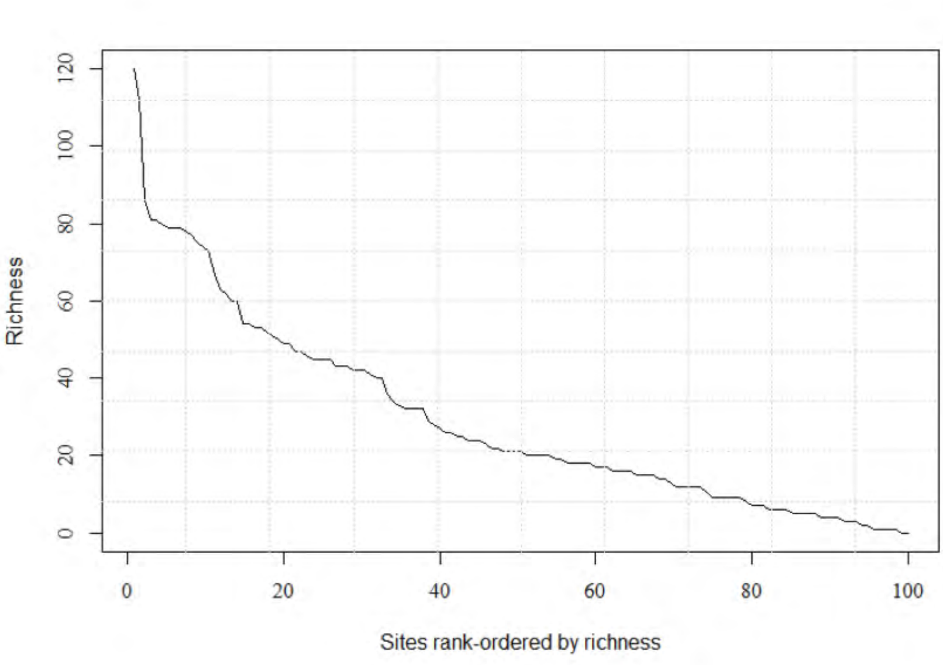
Indicator species in each assemblage were defined by the model using the module of “Indval”. For a given species and a given group of sites, the indicator value is defined as the product of the mean species abundance occurring in the group divided by the sum of the mean abundances in all other groups (a type of specificity), multiplied by the proportion of sites within the group where the species occurs (fidelity) (Dufrene & Legendre, 1997). Each species was associated with the tree node (community) where its maximum indicator value occurred: species with high indicator values were used as characteristic members of each community, and the spatial extent of the group indicated the region where the species was predominantly found (DeVantier *et al.*, 2006).

Data analysis was all performed in R program (R Core Team 2013) with different packages.

Results

Comprehensive description of fish composition

A total of 425 species and subspecies from 15 orders, 42 families had been recorded in the 135 studied lakes. Rank-order species richness is sigmoid (Fig.1a): around 5% of the sites had high species richness (species richness >80), 45% contained moderate richness (20<species richness < 80), and 50% with low richness (species richness< 20). While Fig.1b shows that most of the recorded species are uncommon or rare species, about 80% of the species have an occurrence of less than 20% in all the sites, among which around 47% of the species occurred only once. About 8% of the species are recorded as moderately common species that occurred in more than 30% of the sites. Only 8 species occurred in more than 80 sites and are thus recorded as the most abundant species in our research (*Pseudorasbora parva* (Temminck et Schegel, *Hypophthalmichthys molitrix* (Cuvier et Valenciennes), *Ctenophaxyngodon idellus* (Cuvier et Valenciennes), *Aristichthys nobilis* (Richardson), *Carassius auratus* Linnaeus, *Cyprinus(cyprinus) carpio* Linnaeus, *Misgurnus anguillicaudatus* (Cantor), *Hemiculter leucisculus* (Basilewsky)).



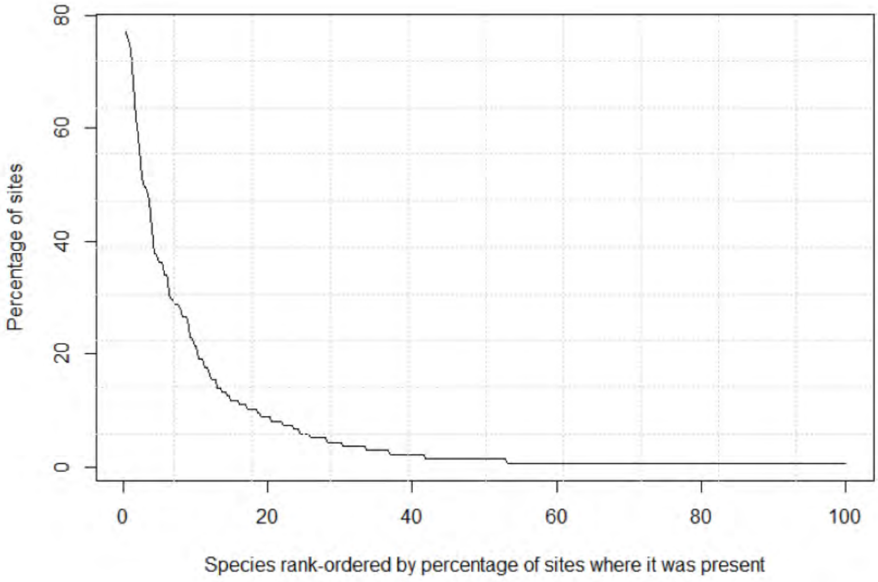


Fig. 1 Rank-ordered distribution plots of fishes in Chinese lakes

a Sites rank-ordered by species richness in each site (X axis represent the % of the 135sites; Y axis represent the richness per site);

b Species (percentage) rank-ordered by percentage of sites where each species encountered (X axis represent the % of 435 species; Y axis represents the total percentage of sites where the species present)

Fish diversity and assemblages in Chinese lakes

In the MRT model, a robust decision tree with five terminal nodes was identified to minimize the cross-validated relative error and the complexity instead of the “1-se rule” as suggested by Breiman *et al.,* (1984) (Fig. 2a). Comparisons of the constrained clustering from MRT models with similar numbers of unconstrained cluster groups (using K-means clustering and Euclidean distance) showed that the constrained and unconstrained groups are similarly homogeneous, which indicates the species clusters were spatially contiguous and the environmental variables adequately accounted for the species variance (De’ath, 2002). Although the MRT tree only accounted for 35.2% of the variation in the species data (Fig. 2b), De’ath described it as common because of the large number of low occurring species (Devantier *et al.,* 2006).

Therefore, all the 135 lakes were patterned into five assemblages according to the similarity of fish composition and the determinants (Fig. 2b). Two main clusters were defined as the first split (Cluster **I** and Cluster **II**), and this split identified two regions: high altitude region (plateau) and low altitude region (plain). In addition, five sub-clusters which we considered as five assemblages were also sub-divided, and contained 61, 26, 38, 4, 6 lakes respectively, named **I**a, **I**b, **II**a1, **II**a2, **II**b.

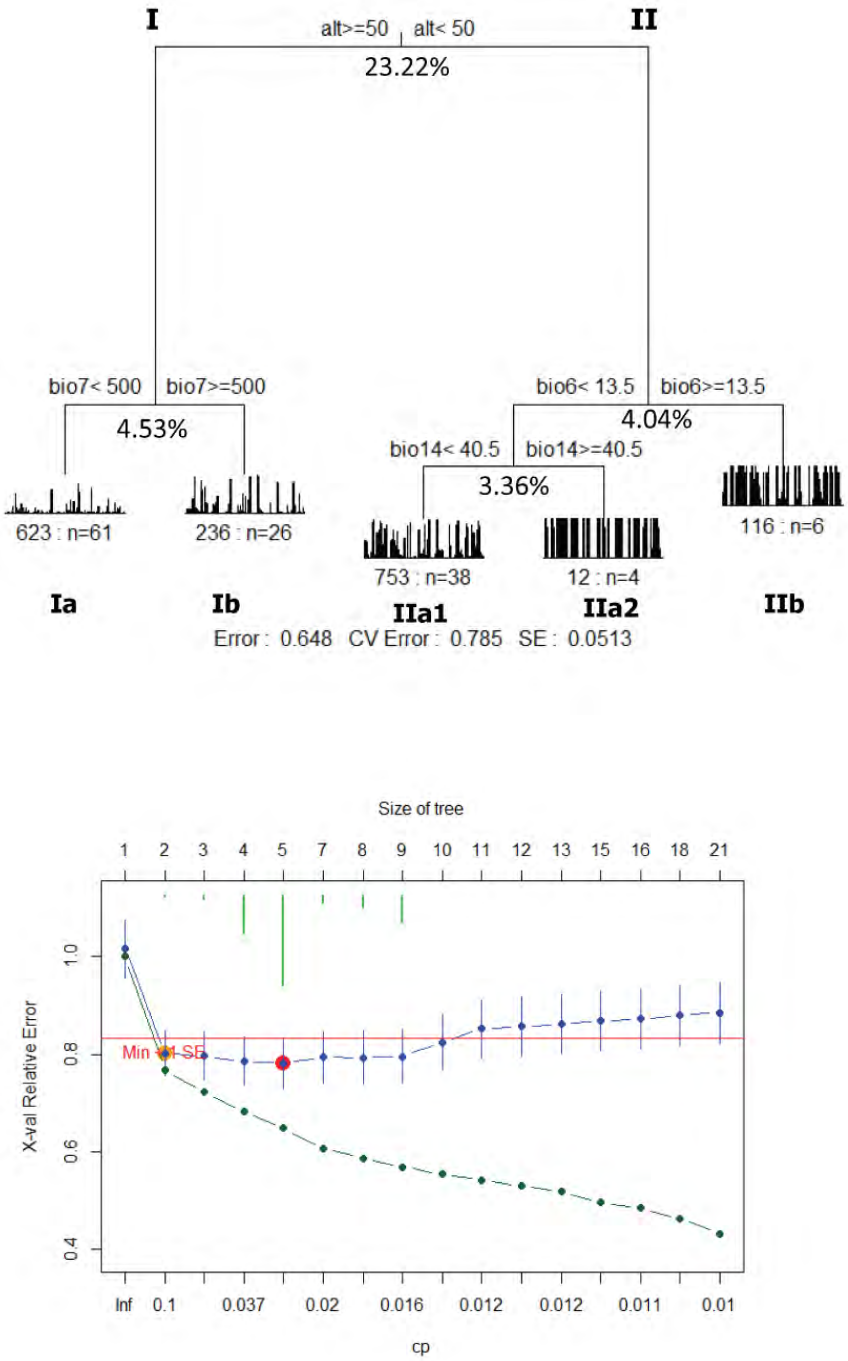


Fig. 2 The output of MRT model for the fish species and environmental variables

a Fish distribution patterns and assemblages defined by MRT model (alt: Altitude of the site; bio6: Minimum temperature of coldest month; bio7: Temperature annual range; bio14: Precipitation of driest month; the terminal nodes represent five fish assemblages **Ia, Ib, IIa1, IIa2, IIb,** see text for details)

b Cross-validation of the Multi-variable Regression Tree analysis (the upper line is cross-validation relative error while the lower line is re-substitution error, the red circled point is the model with the greatest cross-validated predictive accuracy)

Box-plots of species richness in each assemblage are shown in Fig.3. The Kruskall-Wallis test revealed that species richness varied significantly among the different assemblages ($p<0.05$), while the multiple comparison test showed that species richness in each of the three assemblages in cluster **II** were significantly different from the two assemblages in cluster **I** ($p<0.05$). However, there is no significant difference between each assemblage in the same cluster. To simplify, we defined the richness in each lake as “lake richness” while the diversity in the assemblage as “fauna diversity” hereinafter. Cumulative species richness showed that only **I**a did not reach an asymptote compared to the other four (Fig. 4). However, **I**a contained the highest rates of species’ accumulation as well as the highest fauna diversity (280 species; abbreviated as sp. hereinafter), although the species richness in each lake was the lowest (median=12 sp.; Fig.3 and Fig.4). Lake richness in **II**a2 and **II**b showed higher values (median values are 84 sp. and 88.5 sp. respectively) but with lowest accumulative rates, while **II**a2 also contained the lowest fauna diversity with only 84 species. Assemblage **II**a1 had moderate lake richness (median value = 55 sp.) as well as for the fauna diversity (184 sp.). Unfortunately, **I**b seemed to be the one that contained both lower lake richness (median value = 22.5 sp.) and fauna diversity (99sp.) than the other four assemblages (Fig.3 and Fig.4).

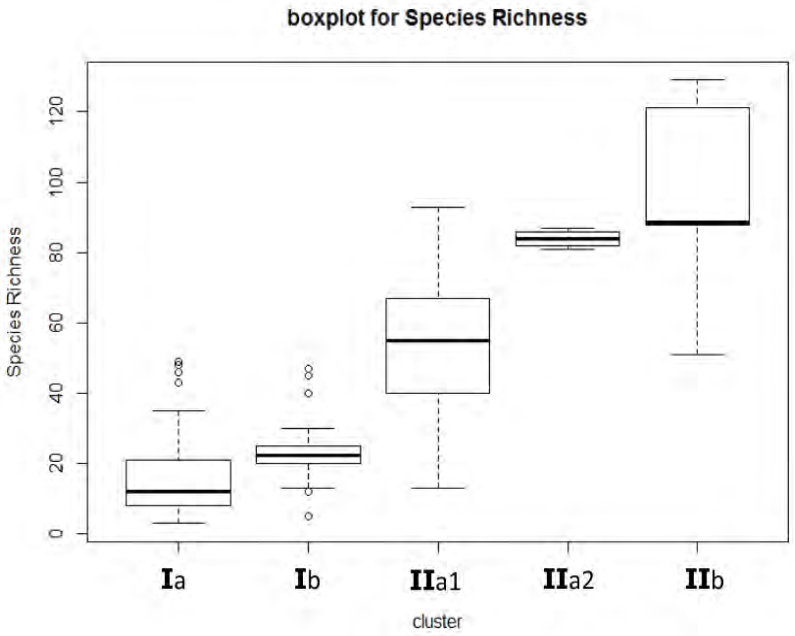


Fig. 3 Box-plot for the species richness of each assemblage (Ia, Ib, IIa1, IIa2, IIb were the five assemblages defined in Fig.3; for each box plot hereinafter in the paper, the median (line within the box), first and third quartiles (box), non-outlier range (whiskers), and outliers (dot) are shown).

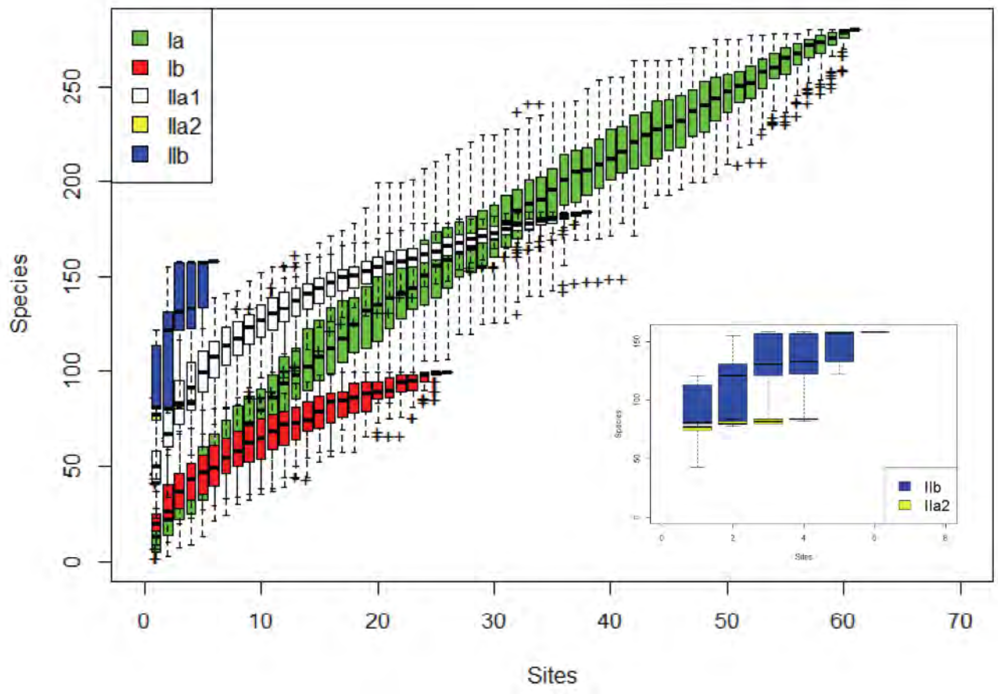


Fig. 4 Species accumulation curves for the five assemblages (figure **a** shows the global picture of the species accumulation for all the five assemblages, while the small figure **b** show the precise details of IIa2, IIb, not clearly shown in figure **a**.)

Determinants of fish distribution and assemblages in Chinese lakes

Five assemblages were thus mapped (Fig. 5), geographically: cluster **I** (**I**a and **I**b) mostly indicated the lakes distributed in Meng-Xin plateau, Yun-Gui plateau, Qinghai-Tibet plateau and the North-East plain, while cluster Ⅱ (Ⅱ a1, Ⅱ a2, Ⅱ b) mainly indicated the lakes distributed in the Eastern plain, with the majority along the Yangtze River basin. From the tree (Fig.2b), we can see that altitude accounted for most of the variation in species distribution (23.22%), and explained about 65.9% of the total variation. Followed by the annual temperature range (bio7; accounted for 4.54%), the minimum temperature of the coldest month (bio6; accounted for 4.04%) and precipitation in the driest month (bio 14; accounted for 3.36%).

For these four determinants, **Ia** was characterized by high altitude (1908.85 ± 1050.96 m) and lower annual temperature range (32.58±10.54℃), while **Ib** in terms of relative high altitude (220.15±202.38m) and higher annual temperature range (51.54±1.47℃). However, cluster Ⅱ, Ⅱ a1 was characterized mostly by the lower altitude (19.45±19.25m), a lower minimum temperature of the coldest month (-0.71±2.62℃) and lower precipitation in the driest month (29.97±9.28mm), while Ⅱ**a2** was characterized by lower altitude (14±0.82m), lower minimum temperature in the coldest month (1.12±0.12℃) and higher precipitation in the driest month (42.75±1.5mm). Assemblage Ⅱ **a2** was divided from assemblage Ⅱ a1 mainly depending on the higher precipitation of the driest month. In Ⅱ **b**, lakes were determined in terms of lower altitude (21.16±5.84m) and higher minimum temperature of the coldest month (2.1±3.84℃).

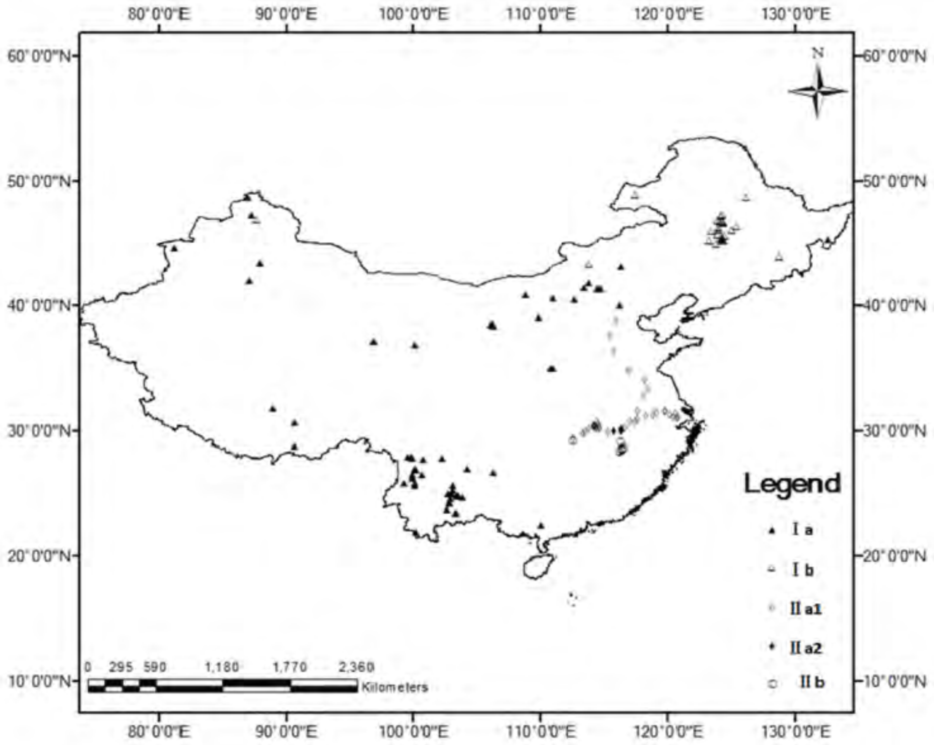


Fig. 5 Map of the five assemblages displayed in mainland China

Indicator species of assemblages

A total of 107 indicator species were indentified in five assemblages (contained 1 sp., 8 sp., 1 sp., 59 sp. and 38 sp. respectively) (see Table. 2 for details). *Cyprinus (cyprinus) chilia* (Wu *et al.*) is the only indicator species for **Ia**, since this assemblage mostly consisted of the plateau fish fauna: fish distributed in these lakes were mostly stenochoric species, with some species even occurring only once. The main indicator species in assemblage **Ib** were *Rhodeus sericeus* (Pallas), *Sarcocheilichthys lacustris* (Dybowsky), *Ladislavia taczanowski* (Dybowsky), *Gobio lingyuanensis* (Mori), *Gobio gobio cynocephalus* (Dybowsky), *Carassius auratus gibelio* (Bloch), *Misgurnus mohoity* (Dybowski), *Perccottus glohni* (Dybowski). *Coilia nasus* (Schlegel) was the indicator species of **Ⅱa1**. Notably, cluster **Ⅱa2** and **Ⅱb** comprised only 10 of the 135 lakes but contained 92.3% of the total indicator species, with all the indicator species in these two assemblages being common species in the shallow lakes along the middle-lower Yangtze River. Therefore, **I**a contains most of the plateau species, **Ⅱa1**, **Ⅱa2** and **Ⅱb** contained mostly the plain fish, while **I**b seems to be the transition zone between plateau and plain as it contained both plain fish and plateau fish.

Discussion

The present study is the first to have assessed the patterns of fish species distribution and assemblages as well as their determinants in a large scale in China. Our research provided precisely understanding of fish diversity and distribution patterns, clarified the main drivers, identified the indicator fish species, verified the hypothesis and examined the capacities of MRT model. Rather than some single models that assume species are independent, MRT is a novel tree method that directly determines the assemblages in terms of environment variables. This is suitable and essential for community analysis and produces more creditable results (De'ath, 2002).

Fish diversity and assemblages in Chinese lakes

In the present study, five assemblages (**Ia**, **Ib**,**Ⅱa1**,**Ⅱa2**,**Ⅱb**) were ultimately redefined according to the determinants and the fish composition, distinguished from the previous studies: nine administrative regions (Kang *et al.*, 2013) and the five physiographic regions (Zhao *et al.*, 2006; Wang & Dou, 1998). Both the lake richness and fauna diversity showed significant differences in each assemblage: lake richness in the plateaus was clearly lower than the plains, while the diversity of the fauna was the highest.

Assemblage **Ⅰa** mainly consisted of the lakes distributed in the three famous plateaus (Qinghai-Tibet plateau, Yun-Gui plateau, Meng-Xin plateau) in China. Due to the particular natural environments and climatic conditions in plateaus (Wang & Dou, 1998), only one indicator species was found since most are stenochoric species. However, the number of lakes contained in this assemblage was as high as 61. This may partly explain why the lake richness in this cluster was lower (median value = 12sp.) but the fauna diversity was the highest (280 sp.). On the other hand, fish fauna diversity in the assemblage will be even higher with the increasing of research lakes since the species accumulation curve still has not approached the asymptote (Fig. 4). Meanwhile, in these plateaus, the complex of river systems and geographic situation, the specific climate and less effect of human activity made these areas easier for energy to be obtained, and thus could support a high number of species (Wang & Dou 1998; Yang et al., 2004; Zhao et al., 2006; Kang et al., 2013). Zhao et al. (2006) also argued that the low lake richness in plateau was the result of the young geological age for new species colonization and speciation. In the authors' opinion, the results sustained the plausibility of the species-energy hypothesis.

Lakes contained in assemblage Ⅰ**b** were mainly distributed in North-east China. In fact, the lakes in this assemblage mainly belong to the Song-Nen plain. However, compared with the low altitude in clusterⅡ，Ⅰ**b** was also divided as high altitude area (altitude>=50m). Indicator species found in Ⅰ**b** were all especially distributed in northern China water systems with low economic value (Yang *et al.*, 2010). A few previous reports considered that although the weather in north-east China was very cold in winter, the fauna richness in this region is not so low (Zhang, 1999; Zhao *et al.*, 2006). Whereas the fauna diversity (99sp.) was even lower than other assemblages in our study, so was the lake richness (median=22.5sp.). Our conclusion was supported by many field investigations (Ren, 1994; Yang *et al.*, 2010), which indicated that the fish diversity in north-east China was significantly lower than in other areas. Overall, assemblages Ⅰ**b** seems to be the transition region from the plateau to the plain since clusterⅡall contained the lakes in the plains with plain fish fauna.

Code	Scientific name	*P*	Asse	Code	Scientific name	*P*	Assem
Cch	*Cyprinus (cyprinus) chilia* Wu et al.	0.029	Ⅰ a	Ssm	*Silurus soldatovi meridioalis* Chen	0.002	Ⅱ a2
Rse	*Rhodeus sericeus* (Pallas)	0.01	Ⅰ b	Cfus	*Clarias fuscus*	0.003	Ⅱ a2
Slac	*Sarcocheilichthys lacustris* (Dybowsky)	0.047	Ⅰ b	Pvac	*Pseudobagrus vachelli* (Richardson)	0.001	Ⅱ a2
Lta	*Ladislavia taczanowski* (Dybowsky)	0.028	Ⅰ b	Ppg	*Pseudobagrus pratti* Gunther	0.001	Ⅱ a2
Gli	*Gobio lingyuanensis* Mori	0.005	Ⅰ b	Lni	*Liobagrus nigricauda* (Regan)	0.002	Ⅱ a2
Ggc	*Gobio gobio cynocephalus* Dybowsky	0.006	Ⅰ b	Aja	*Anguilla japonica* T.et S.	0.006	Ⅱ a2
Cag	*Carassius auratus gibelio* (Bloch)	0.001	Ⅰ b	Ola	*Oryzias latipes* (Schlegel)	0.021	Ⅱ a2
Mmo	*Misgurnus mohoity* (Dybowski)	0.022	Ⅰ b	Sch	*Siniperca scherzeri* Steindachner	0.033	Ⅱ a2
Pglo	*Perccottus glohni* (Dybowski)	0.001	Ⅰ b	Cro	*Coresiniperca roulai* Wu	0.002	Ⅱ a2
Cna	*Coilia nasus (*Schlegel)	0.028	Ⅱ a1	Mch	*Macropodus chinensis* (Bloch)	0.017	Ⅱ a2
Cbr	*Coilia brachygnathus* (Kreyenberg et	0.007	Ⅱ a2	Oob	*Odontobutis obscura* (Temminck et Schlegel)	0.015	Ⅱ a2
Ntt	*Neosalanx tankankei taihuensis* Chen	0.004	Ⅱ a2	Mmy	*Mugilogobius myxodermus* (Herre)	0.001	Ⅱ a2
Nol	*Neosalanx oligodontis* Chen	0.005	Ⅱ a2	Ccl	*Ctenogobius cliffordpopei* (Nichols)	0.002	Ⅱ a2
Eba	*Elopichthys bambusa* (Richardson)	0.012	Ⅱ a2	Cgi	*Ctenogobius giurinus* (Rutter)	0.026	Ⅱ a2
Oel	*Ochetobius elongatus* (Kner)	0.004	Ⅱ a2	Asi	*Acipenser sinensis* Gray	0.01	Ⅱ b
Scu	*Squaliobarbus curriculus* (Richardson)	0.008	Ⅱ a2	Pgl	*Psephurus gladius* (Martens)	0.008	Ⅱ b
Obi	*Opsariichthys bidens* (Gunther)	0.004	Ⅱ a2	Mre	*Macrura reevesi* (Richardson)	0.02	Ⅱ b
Tsw	*Toxabramis swinhonis* (Gunther)	0.01	Ⅱ a2	Cec	*Coilia ectenes* (Jordan et Seale)	0.001	Ⅱ b
Hbw	*Hemiculter bleekeri* Warpachowsky	0.01	Ⅱ a2	Hbr	*Hemisalanx branchyrostralis* (Fang)	0.005	Ⅱ b
Ppek	*Parabramis pekinensis* (Basilewsky)	0.028	Ⅱ a2	Lma	*Luciobrama macrocephalus* (Lacepede)	0.001	Ⅱ b
Psi	*Pseudolaubuca sinensis* (Bleeker)	0.001	Ⅱ a2	Zpl	*Zacco platypus* (Temminck et Schlegel)	0.003	Ⅱ b
Psim	*Pseudobrama simoni*(Bleeker)	0.001	Ⅱ a2	Oub	*Opsariichthys uncirostris bidens* (Gunther)	0.006	Ⅱ b
Mte	*Megalobrama terminalis* (Richardson)	0.015	Ⅱ a2	Tar	*Toxabramis argentifer*	0.001	Ⅱ b
Eda	*Erythroculter dabryi* (Bleeker)	0.008	Ⅱ a2	Parg	*Pseudobagrus argentivitatus*(Regan)	0.001	Ⅱ b
Emo	*Erythroculter mongolicus* (Basilewsky)	0.014	Ⅱ a2	Paen	*Parapelecus engraulis* (Nichols)	0.001	Ⅱ b
Cox	*Culter oxycephalus* (Bleeker)	0.001	Ⅱ a2	Mam	*Megalobrama amblycephala* Yih	0.029	Ⅱ b
Eil	*Erythroculter ilishaeformis* (Bleeker)	0.036	Ⅱ a2	Eox	*Erythroculter oxycephaloides* Kreyenberg et	0.002	Ⅱ b
Xda	*Xenocypris davidi* Bleeker	0.005	Ⅱ a2	Pmi	*Plagiognathops microlepis* (Bleeker)	0.005	Ⅱ b
Xmi	*Xenocypris microlepis* Bleeker	0.001	Ⅱ a2	Dtu	*Distoechodon tumirostris* Peters	0.001	Ⅱ b
Xar	*Xenocypris argentea* (Gunther)	0.008	Ⅱ a2	Asim	*Acanthobrama simoni* (Bleeker)	0.003	Ⅱ b
Dhu	*Distoechodon hupeinensis* (Yih)	0.001	Ⅱ a2	Pli	*Psendoperilampus lighti*	0.001	Ⅱ b
Roc	*Rhodeus ocellatus (Kner)*	0.016	Ⅱ a2	Ahy	*Acheilognathus hypselonotus* Bleeker	0.001	Ⅱ b
Rfa	*Rhodeus fangi* Fang	0.001	Ⅱ a2	Sca	*Spinibarbus caldwelli* (Nichols)	0.001	Ⅱ b
Rli	*Rhodeus Lighti* Wu	0.002	Ⅱ a2	Spa	*Schizopyge parvus* Tsao	0.001	Ⅱ b
Pii	*Paracheilognathus iimberbis* (Gunther)	0.001	Ⅱ a2	Ski	*Sarcocheilichthys kiangsiensis* Nichols	0.001	Ⅱ b
Aba	*Acheilognathus barbatulus* Gunther	0.001	Ⅱ a2	Rcy	*Rhinogobio cylindricus* Gunther	0.001	Ⅱ b
Amac	*Acheilognathus macropterus (*Bleeker)	0.032	Ⅱ a2	Sdu	*Saurogobio dumerili* Bleeker	0.001	Ⅱ b
Ach	*Acheilognathus chankaensis* (Dybowsky)	0.007	Ⅱ a2	Gic	*Gobiobotia ichangensis* Fang	0.003	Ⅱ b

Ato	*Acheilognathus tonkinensis* (Vaillant)	0.003	Ⅱa2	Cma	*Cobitis macrostigma* Dabry	0.001	Ⅱb
Pgu	*Paracanthobrama guichenoti* Bleeker	0.007	Ⅱa2	Pfa	*Parabotia fasciata* Dabry	0.001	Ⅱb
Hma	*Hemibarbus maculatus* Bleeker	0.021	Ⅱa2	Lel	*Leptobotia elongata* (Bleeker)	0.01	Ⅱb
Sss	*Sarcocheilichthys sinensis sinensis* Bleeker	0.006	Ⅱa2	Mda	*Misgurnus dabryanus* Sauvage	0.004	Ⅱb
Snn	*Sarcocheilichthys nigripinnis nigripinnis*	0.014	Ⅱa2	Peu	*Pseudobagrus eupogon* (Boul.)	0.001	Ⅱb
Snit	*Squalidus nitens* (Gunther)	0.001	Ⅱa2	Pal	*Pseudobagrus albomarginatus* (Rendhal)	0.001	Ⅱb
Sar	*Squalidus argentatus* (Sauvage et Dabry)	0.001	Ⅱa2	Llo	*Leiocassis longirostris* Gunther	0.05	Ⅱb
Che	*Coreius heterodon*(Bleeker)	0.001	Ⅱa2	Lcr	*Leiocassis crassilabris* Gunther	0.034	Ⅱb
Rty	*Rhinogobio typus* Bleeker	0.007	Ⅱa2	Lus	*Leiocassis ussuriensis* (Dybowski)	0.046	Ⅱb
Mtu	*Microphysogobio tungtingensis* (Nichols)	0.042	Ⅱa2	Hmac	*Hemibagrus macropterus* Bleeker	0.014	Ⅱb
Mki	*Microphysogobio kiatingensis* (Wu)	0.001	Ⅱa2	Gsi	*Glyptothorax sinensis* (Regan)	0.004	Ⅱb
Sda	*Saurogobio dabryi* Bleeker	0.012	Ⅱa2	Hku	*Hemirhamphus kurumeus* (Jordan et Starks)	0.012	Ⅱb
Sgy	*Saurogobio gymnocheilus Lo* Yao et Chen	0.001	Ⅱa2	Skn	*Siniperca kneri* Garman	0.019	Ⅱb
Csi	*Cobitis sinensis* (Sauvage et Dabry)	0.001	Ⅱa2	Ssc	*Siniperca scherzeri* Steindachner	0.001	Ⅱb
Pba	*Parabotia banarescui* (Nalbant)	0.001	Ⅱa2	Mop	*Macrodus opercularis* (Linnaeus)	0.001	Ⅱb
Lpu	*Leptobotia purpurea* (Nichols)	0.001	Ⅱa2				

Table 2 Indicator species in each assemblage in Chinese lakes

Generally, lakes contained in cluster Ⅱ could roughly be considered as the eastern plain in terms of lower altitude (altitude<50m), coinciding with the classification of Zhang (1954) and Li (1981). Assemblage Ⅱ**a1** consisted of the largest part of eastern region lakes (38 in total of 48) with *Coilia nasus* (Schlegel) as the sole indicator species. Simultaneously, lakes contained in assemblage Ⅱ**a1** contained moderate lake richness (median value = 55 sp.) and fauna diversity (184 sp.) compared with other assemblages. From the map, we can see that the assemblage occupied two parts, some few lakes distributed in the 3H plain region (Huanghe, Huaihe and Haihe) (Kang *et al.*, 2013), but most of them distributed in the middle-lower Yangtze plain region. The superiority of the natural environment along with the abundance of lake resources has turned the lakes in this area into one of the most important freshwater fishery bases in central China (Cui & Li, 2005).

AssemblageⅡ**a2** only contained 4 lakes that were also distributed in the middle Yangtze River basin. The proper climatic conditions along with ample natural resources in the middle Yangtze River basin showed advantages for supporting more fish species (Liang & Liu, 1995; Cui & Li, 2005; Zhang, 2005), and also concurred with the species-energy hypothesis (Wright, 1983). 59 indicator species were identified in the assemblage. All the indicator species were common and dominant species in the lakes along the middle-lower Yangtze River basin. Assemblage Ⅱb consisted of 6 lakes distributed along the central Yangtze River basin, which contained the highest lake richness (with median value of 88.5 sp.) among all five assemblages. 38 indicator species were thus identified in the assemblage, similar to assemblage Ⅱ**a2**, the indicator species in assemblage Ⅱ**b** were also common plain species in the central-lower Yangtze River. In fact, lakes contained in assemblage Ⅱ**a2** and Ⅱ**b** were either connected to the Yangtze River (*e.g.* Dongting lake, Poyang lake, Junshan lake, Wuhu lake) or isolated from some large lakes (*e.g.* Huanghu lake and Huangdahu lake). Moreover, these lakes used to be connected to the Yangtze River (Shi, 1989) which supported even more common fish species (Fu *et al.*, 2003). Due to the historical river-lake connection, fish composition showed highly homogeneity in these lakes (Ⅱa2 and Ⅱb, no significant difference), and therefore contained most of the indicative species (87.8% of all the indicator species). Therefore, it is no wonder that lakes distributed along the Yangtze River basin have higher lake richness compared with other assemblages. This result was also consistent with recent studies (Zhao *et al.,* 2006; Kang *et al.,* 2013). Consequently, compared with **Ⅱa2** and **Ⅱb**, **Ⅱa1** demonstrated lower lake richness (median value = 55 sp.) mostly because of the disconnection of the lakes and the river (Liu & Wang, 2010; Fang *et al.,* 2006). Furthermore, lake degradation, environmental pollution and irrational fishing modes in these lakes can also decrease the fish richness (Cui & Li, 2005; Fang *et al.,* 2006).

Determinants of lake fish distribution and assemblages

Overall, altitude, temperature annual range (TAN), precipitation in the driest month (PDM) and the minimum temperature of the coldest month (MTCM) were ultimately identified as the key factors for determining fish distribution and assemblages in Chinese lakes, among which altitude was the most important determinant. Altitude, precipitation and temperature have long been acknowledged to determine the distribution of organisms in various ecosystems, such as plants (Pottier *et al.*, 2013), riverine and marine fish and invertebrates (Bussion *et al.*, 2008; Cheung *et al.*, 2009).

Our result fitted very well with previous research that showed that altitude considerably affects fish distribution in aquatic ecosystems (Amarasinghe & Welcomme, 2002; Yoon *et al.,* 2011; Barradas *et al.,* 2012; Stojkovic *et al.,* 2013). However, altitude is a complex variable: it can cause direct and indirect effects on fish distribution. Some preview studies have explored the two different influences (Legendre, 1993; Hawkins *et al.,* 2003; Zhao *et al.,* 2006), and have shown that indirect influences were even greater than direct effects, since altitude contained many other factors that could affect fish distribution.

Temperature was an important factor that influenced fish distribution through its affect on fish metabolism (Gillooly *et al.,* 2001), breeding (Mills & Mann, 1985), development and growth (Mann, 1996; Wolter, 2007) and behavior (Taniguchi *et al.*, 1998). Here in our study, two kinds of temperature (TAN and MTCM) have been defined as the important factors which determine the fish distribution, consistent with Crisp (1996) and Mann (1996) who considered these two factors as having great importance in freshwater fish biological requirements. TAN heavily contributed in dividing cluster Ⅰ into two parts, the thermal range appeared to be important in the present study partly because of its vast variability (21.2℃-56.1℃). Since fish are poikilothermal animals they will suffer considerably from the environment temperature change as demonstrated on some European freshwater fish (Buisson *et al.*, 2008; Pont *et al.*, 2005) and American species (Rathert *et al.*, 1999). Although each fish species has a thermal preference and can usually be characterized by a tolerance range of temperatures varying by a few degrees, they are flexible and can adapt to a certain extent by physiological adjustments or behavioral thermoregulation. However, thermal ranges that exceed the proper range will without doubt affect fish lives and distributions (Gislason *et al.*, 2010). This is the explanation as to why the MTCM was a major determinant of the fish distribution and assemblages in our study. In the present study, MTCM acted as the main determinants in cluster Ⅱ. This may be mainly because the fishes distributed in assemblage Ⅱ were adapted to temperate and a subtropical zone climate. Globally, MTCM had also been defined as the most important factor that determines the fish distributions and lives (Rubidge *et al.*, 2011; Yu *et al.*, 2013; Aguilar-Kirigin & Naya, 2013). Other than these two factors, precipitation has always been considered as one of the most important climatic factors in numerous recent studies (Zhao *et al.*, 2006; Buisson *et al.*, 2008; Buisson & Grenouillet, 2009), because the precipitation could influence the stream flows and hydrological conditions. The hydrology was even more complex and concerns diverse fields such as meteorology, geomorphology, geology or geography which could all affect fish assemblages and distribution. The PDM was also considered as having major impacts on niche shifts of freshwater species (Lauzeral *et al.*, 2011), and thus affect fish distribution. Indeed, the PDM modifies the hydrological conditions of lakes directly, such as water content, water depth and transparency, and nutrients (Wetzel, 2001), which in turn affect fish lives and distributions (Brazner & Beals, 1997; Petry *et al.*, 2003; Mello *et al.*, 2009; Kang *et al.*, 2013). Indeed, besides these factors, there are still many others that could affect and hinder fish distribution and biodiversity, such as land-cover, geographical variables, topological variables, biological invasion, artificial impacts and even net primary production (Park *et al.*, 2006; Buisson *et al.*, 2008; Gevrey *et al.*, 2009; Maloney *et al.*, 2013; Kang *et al.*, 2013). Thus in future studies, more impact factors should be considered comprehensively in order to obtain more powerful results and deeper explanations about the fish distributions and assemblages in lakes.

In conclusion, altitude, precipitation and temperature which could also be attributed to the energy availability explain the fish diversity and distribution patterns across China. However, lake area that have always been considered as the key factor to determine fish dispersal, failed to explain the patterns.All the findings in the current study interpreted that fish species distribution pattern seems to support the species-energy hypothesis rather than the species-area hypothesis.

Potential implications for biodiversity conservation

Our results provide useful perspectives, not only on the current distribution of fish species, but also on the future scenarios under global change.These results revealed important perspectives on species conservation as well as the management of aquatic ecosystems in the future, and thus will be of interest to conservation biologists and environment managers.

Overall, in terms of our findings, the three plateaus that comprised cluster **Ⅰa** should receive high conservation priority on fish conservation since **Ⅰa** possessed of the highest fauna diversity and richness of endemic fish species. Potential natural reserves should be considered for the conservation of endemic and endangered fish species. Indeed, the construction of natural reserves has long been considered as one of the effective approaches to conserve fish resources (Park *et al.*, 2003; 2004; Zhao *et al.*, 2006; Kang *et al.*, 2013). These conservation strategies are mostly focused on the endemic fish since they are abundant in Chinese lakes and rivers (Park *et al.*, 2003;

Fu *et al.,* 2004; He *et al.,* 2011). However, we still cannot ignore the impacts of invasive species since they are considered as one of the greatest threats to global biodiversity (Butchart *et al.,* 2010; Vorosmarty *et al.,* 2010). This is especially true for freshwater ecosystems that are particularly vulnerable to biological invasions (Cucherousset & Olden, 2011).

Otherwise, in the authors' opinion, the most important plain region, **Ⅱa1,** that contains most of the plain fish fauna together with the highest lake richness needs extra consideration, since **Ⅱa1** contains most of the lakes which provide major commercial freshwater fisheries in central China (Cui & Li, 2005). As conservation should be carried out in a networked region rather than a single reserve (Bonn & Gaston, 2005), different conservation strategies should be proposed aimed at different objectives. By contrast, with **Ⅰa**, the conservation strategies here should be to upgrade the fisheries to an eco-friendly modes, to improve the fish diversity and the health of the ecosystem (Guo *et al.,* 2013). Moreover, one suggestion for the conservation of fish diversity in this area was the connection of the lakes with the Yangtze River (Fu *et al.,* 2003; Liu *et al.,* 2010).

Acknowledgements

Grateful thanks are due to Prof. Gael Grenouillet, Dr. Mathieu Chevalier, Dr. Loic Tudesque for their technical support. Thanks John Woodley for correcting English. This research was financially supported by Grant Nos. 200903048 and 20130305 from the “Special Fund for Agro-scientific Research in the Public Interest” of China, Nos. 30830025 and 30900182 from the National Natural Science Foundation of China.

References

Aguilar-Kirigin Á. J. & Naya D. E. (2013) Latitudinal patterns in phenotypic plasticity: the case of seasonal flexibility in lizards’ fat body size. *Oecologia*, **173**,745-752.

Amarasinghe U. S. & Welcomme R. L. (2002) An analysis of fish species richness in natural lakes. *Environmental Biology of Fishes,* **65**, 327-339.

Araújo M.B. & Guisan A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677–1688.

Barradas J. R. S., Silva L. G., Harvey, B. C. & Fontoura, N. F. (2012) Estimating migratory fish distribution from altitude and basin area: a case study in a large Neotropical river. *Freshwater Biology*, **57**, 2297-2305.

Brazner J.& Beals E. (1997) Patterns in fish assemblages from coastal wetland and beach habitats in Green Bay, Lake Michigan: a multivariate analysis of abiotic and biotic forcing factors. *Canadian Journal of Fishery Aquatic Science*,**54**, 1743－1761.

Breiman L., Friedman J. H., Olshen R. A. & Stone C. G. (1984) *Classification and regression trees*. Wadsworth International Group, Belmont, California, USA.

Brummer T. J., Maxwell B. D., Higgs M. D. & Rew L. J. (2013) Implementing and interpreting local－scale invasive species distribution models. *Diversity and Distributions*, **19**, 919-932.

Buisson L. & Grenouillet G. (2009) Contrasted impacts of climate change on stream fish assemblages along an environmental gradient. *Diversity and Distributions*, **15**, 613–626.

Buisson L., Thuiller W., Lek S., Lim P. & Grenouillet G. (2008) Climate change hastens the turnover of stream fish assemblages. *Global Change Biology*, **14**, 2232–2248.

Butchart S. H., Walpole M., Collen B., van Strien A., Scharlemann J. P. et al. (2010) Global Biodiversity: Indicators of Recent Declines. *Science*, **28**, 1164-1168

Cawsey E.M., Austin M.P. & Baker B.L. (2002) Regional vegetation mapping in Australia: a case study in the practical use of statistical modelling. *Biodiversity and Conservation*, **11**, 2239–2274.

Cheung W. W., Lam V. W., Sarmiento J. L., Kearney K., Watson R. & Pauly D. (2009) Projecting global marine biodiversity impacts under climate change scenarios. *Fish and Fisheries*, **10**, 235-251.

Crisp D.T. (1996) Environmental requirements of common riverine European salmonid fish species in freshwater with particular reference to physical and chemical aspects. *Hydrobiologia*, **323**, 201－221.

Cui Y.B. & Li Z.J. (2005) *Fishery Resources and Conservation of Environment in Lakes of the Changjiang River Basin*. Science Press, Beijing (in Chinese).

Cucherousset J.& Olden D. (2011) Ecological Impacts of Nonnative Freshwater Fishes. *Fisheries,***36**, , 215-230

Davidson T. A., Sayer C. D., Perrow M., Bramm M. & Jeppesen E. (2010) The simultaneous inference of zooplanktivorous fish and macrophyte density from sub−fossil cladoceran assemblages: a multivariate regression tree approach. *Freshwater Biology*, **55**, 546-564.

De'ath G. (2002) Multivariate regression trees: a new technique for modeling species-environment relationships. *Ecology*, **83**, 1105-1117.

De'ath G. & Fabricius K. E. (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*, **81**, 3178-3192.

DeFries R., Foley J. & Asner G. (2004) Land-use choices: balancing human needs and ecosystem function. *Frontiers in Ecology and the Environment*, **2**,249–57.

DeVantier L. M., De’Ath G., Turak E., Done T. J. & Fabricius K. E. (2006) Species richness and community structure of reef-building corals on the nearshore Great Barrier Reef. *Coral Reefs*, **25**, 329-340.

Dufrene M.& Legendre P. (1997) Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monograph,* **67**, 345－366.

Elith, J. & Leathwick J. (2007) Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and Distributions*,**13**,265－275.

Fang J.Y., Wang Z.H., Zhao S.Q., Li Y.K. et al. (2006) Biodiversity changes in the lakes of the Central Yangtze. *Frontiers in Ecology and the Environment,* **4**, 369-377.

Fu C.Z., Wu J.H., Chen J.K., Wu Q.H. & Lei G.C. (2003) Freshwater fish biodiversity in the Yangtze River basin of China: patterns, threats and conservation. *Biodiversity and Conservation*, **12**, 1649-1685.

Gislason H., Daan N., Rice J. C. & Pope J. G. (2010) Size, growth, temperature and the natural mortality of marine fish. *Fish and Fisheries*, **11**: 149–158.

Gillooly J.F., Brown J.H., West G.B., Savage Van M. & Charnov E.L. (2001) Effects of size and temperature on metabolic rate. *Science*, **293**, 2248 – 2251.

Giraudoux P., Pleydell D., Raoul F., Quéré J. P. et al. (2006). Transmission ecology of *Echinococcus multilocularis*: What are the ranges of parasite stability among various host communities in China? *Parasitology international*, **55**,237-246.

Grenouillet G., Buisson L., Casajus N. & Lek S. (2011). Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography*, **34**: 9-17.

Guisan A. & Zimmerman N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147-186.

Guo C.B.,Ye S.W., Lek S., Liu J.S., Zhang T.L.& Li Z.J. The need for improved fishery management in a shallow macrophytic lake in the Yangtze River basin: evidence from the food web structure and ecosystem analysis. *Ecological Modelling*, DOI: 10.1016/j.ecolmodel.2013.07.013

Hamann A., Gylander T. & Chen P. Y. (2011) Developing seed zones and transfer guidelines with multivariate regression trees. *Tree Genetics & Genomes*, **7**, 399-408.

Hawkins B. A., Field R., Cornell H. V., Currie D. J. et al.(2003). Energy, water, and broad-scale geographic patterns of species richness. *Ecology*, **84**, 3105-3117.

He Y.F.,Wang J.W., Lek S., Cao W.X. & Lek-ang S.(2011) Structure of endemic fish assemblages in the upper Yangtze River basin. *River research and application*, **27**, 59-75.

Hijmans R. J., Cameron S. E., Parra J. L., Jones P. G. & Jarvis A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*,**25**, 1965–1978.

Hutchinson M. F. (1995) Interpolating mean rainfall using thin plate smoothing splines. *International Journal of geographical information systems*, **9**, 385-403.

Kang B., Deng J., Wu Y., Chen L., Zhang J., Qiu H., Lu Y. & He D. (2013) Mapping China's freshwater fishes: diversity and biogeography. *Fish and Fisheries*. doi:

10.1111/faf.12011

Lauzeral C., Leprieur F., Beauchard O., Duron Q., Oberdorff T., & Brosse S. (2011) Identifying climatic niche shifts using coarse-grained occurrence data: a test with non‐native freshwater fish. *Global Ecology and Biogeography*, **20**, 407-414.

Legendre P. (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology*, **74**, 1659-1673.

Li S.Z. (1981) *Studies on Zoogeographical Divisions for Freshwater Fishes in China*. Science Press, Beijing (in Chinese).

Liang Y.L. & Liu H.Q. (1995) *Resources, environment and fishery ecological management of macrophytic lakes*. The Science Press, Beijing (in Chinese with English abstract).

Liu X., & Wang H. (2010) Estimation of minimum area requirement of river‐connected lakes for fish diversity conservation in the Yangtze River floodplain. *Diversity and Distributions*, **16**, 932-940.

Magurran A.E., (2009) Threats to Freshwater Fish. *Science*, **325**,1215-1216.

Maloney K. O. , Weller D. E. , Michaelson D. E. & Ciccotto P. J.(2013) Species Distribution Models of Freshwater Stream Fishes in Maryland and Their Implications for Management. *Environmental Modeling & Assessment* , **18**:1-12. DOI 10.1007/s10666-012-9325-3

Mann R.H.K. (1996) Environmental requirements of European non-salmonid fish in rivers. *Hydrobiologia* , **323**, 223-235.

Mello T. D., Meerhoff M., PEKCAN‐HEKIM, Z. E. Y. N. E. P., & Jeppesen, E. (2009) Substantial differences in littoral fish community structure and dynamics in subtropical and temperate shallow lakes. *Freshwater Biology*, **54**, 1202-1215.

Mills C.A. & Mann R.H.K. (1985) Environmentally-induced fluctuations in year-class strength and their implications for management. *Journal of Fish Biology,* **27**, 209–226.

Ouellette M. H., Legendre P. & Borcard D. (2012) Cascade multivariate regression tree: a novel approach for modelling nested explanatory sets. *Methods in Ecology and Evolution*, **3**, 234-244. Park Y-S, Chang J.B., Lek S., Cao W.X., Brosse S.

(2003) Conservation strategies for endemic fish species threatened by the three Gorges Dam. *Conservation Biology,* **17**: 1748-1758.

Park Y-S, Grenouillet G., Esperance B., Lek S. (2006) Stream fish assemblages and basin land cover in a river network. *Science of the Total Environment*, **365**, 140-153.

Petry P., Bayley P. & Markle D. (2003) Relationships between fish assemblages, macrophytes and environmental gradients in the Amazon River floodplain. *Journal of Fish Biology*, **63**, 547-579.

Pont D., Hugueny B. & Oberdorff T. (2005) Modelling habitat requirement of European fishes: do species have similar responses to local and regional constraints? *Canadian Journal of Fisheries and Aquatic Sciences,* **62**, 163-173.

Pottier J., Dubuis A., Pellissier L., Maiorano L., Rossier L., Randin C. F., Vittoz P., Guisan A. (2013) The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. *Global Ecology and Biogeography,* **22**: 52–63.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rathert D., White D., Sifneos J.C. & Hughes R.M. (1999) Environmental correlates of species richness for native freshwater fish in Oregon, USA. *Journal of Biogeography,* **26**, 257 – 273.

Ren M.L. (1994) Fish fauna of Heilongjiang. *Chinese Journal of Fisheries,***1**:1-14.

Roura-Pascual N., Brotons L., Peterson A. T. & Thuiller W. (2009). Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. *Biological Invasions*, **11**, 1017-1031.

Rubidge E. M., Monanan W. B., Parra J. L., Cameron S. E. & Brashares J. S. (2011). The role of climate, habitat, and species cooccurrence as drivers of change in small mammal distributions over the past century. *Global Change Biology*, **17**, 696-708.

Rushton S.P., Ormerod S.J. & Kerby G. (2004) New paradigms for modelling species distributions? *Journal of applied ecology*, **41**, 193-200.

Sharma S., Jackson D. A., Minns C. K. & Shuter B. J. (2007) Will northern fish populations be in hot water because of climate change? *Global Change Biology*, **13**, 2052–2064.

Shi C.Y. (1989) *A general outline of Chinese lakes*. Science Press, Beijing (in Chinese).

Stojkovic M., Simic V., Milosevic D., Mancev D., & Penczak T. (2013) Visualization of fish community distribution patterns using the self-organizing map: A case study of the Great Morava River system (Serbia). *Ecological Modelling*, **248**, 20-29.

Taniguchi Y., Rahel F.J., Novingen D.C. & Gerow K.G.(1998) Temperature mediation of competitive interactions among three fish species that replace each other along longitudinal stream gradients. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 1894-1901.

Thuiller W., Lavorel S., Araújo M. B., Sykes M. T., & Prentice I. C. (2005) Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 8245-8250.

Tisseuil C., Leprieur F., Grenouillet G., Vrac M., & Lek S. (2012) Projected impacts of climate change on spatio‐temporal patterns of freshwater fish beta diversity: a deconstructing approach. *Global Ecology and Biogeography*, **21**, 1213-1222.

Vayssières M. P., Plant R. E. & Allen‐Diaz B. H. (2000) Classification trees: An alternative non‐parametric approach for predicting species distributions. *Journal of vegetation science*, **11**, 679-694.

Vörösmarty C. J., McIntyre P. B., Gessner M. O., Dudgeon D., Prusevich A., Green P., Glidden S., Bunn S. E., Sullivan C. A., Reidy Liermann C. & Davies P. M. (2010) Global threats to human water security and river biodiversity, *Nature*, **467**,555–561.

Wang S.M., Dou H.S. (1998) *Lakes of China*. Science Press, Beijing.

Wetzel R. G. (2001) *Limnology: lake and river ecosystems.* Academic press.

Wolter C. (2007) Temperature influence on the fish assemblage structure in a large lowland river, the lower Oder River, Germany. *Ecology of freshwater fish*, **16**, 493-503.

Wright D.H. (1983) Species–energy theory: an extension of species area theory. *Oikos*, **41**, 496-506.

Wu X.W. (1964) *The Cyprinoid Fishes of China,* Vol. 1. People’s Press, Beijing (in Chinese).

Wu X.W. (1977) *The Cyprinoid Fishes of China,* Vol. 2. People’s Press, Beijing (in Chinese).

Xie S.G., Cui Y.B. & Li Z.J. (2001) Small fish communities in two regions of the Liangzihu Lake, China, with or without submersed macrophytes. *Journal of Applied Ichthyology*. **17**, 89-92.

Yang F.，Lu X.，Lou Y.，Xue B.，Yao S.，Wei X.，Li Z. & Shan Y. (2010) Structures of fish assemblages from lakes in Songnen Plain. *Journal of Lake Science*, **22**,842-851(in Chinese with English abstract).

Yang Y.M., Tian K., Hao J.M., Pei S.J. & Yang Y.X. (2004) Biodiversity and biodiversity conservation in Yunnan, China. *Biodiversity and Conservation*, **13**, 813-823.

Ye S.W., Li Z., Lek-Ang S., Feng G., Lek S. & Cao W. (2006) Community structure of small fishes in a shallow macrophytic lake (Niushanhu Lake) along the middle reach of the Yangtze River, China. *Aquatic Living Resources,* **19**, 349–359.

Yin M. Ecology of fish [M].China agriculture Press, Beijing, 1982 (In Chinese)

Yoon J. D., Kim J. H., Byeon M. S., Yang H. J. et al. (2011) Distribution patterns of fish communities with respect to environmental gradients in Korean streams. *Annales de Limnologie - International Journal of Limnology,* **47**, S63-S71.

Yu D., Chen M., Zhou Z., Eric R., Tang Q., & Liu H. (2013) Global climate change will severely decrease potential distribution of the East Asian coldwater fish Rhynchocypris oxycephalus (*Actinopterygii, Cyprinidae*). *Hydrobiologia*, **700**, 23-32.

Zhang C.L. (1954) Distribution of Chinese freshwater fishes. *Acta Geographica Sinica* **3**, 279-284(in Chinese).

Zhang T.L. (2005) Life-history strategies, trophic patterns and community structure in the fishes of Lake Biandantang. PhD Dissertation. Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan (in Chinese with English abstract).

Zhang R.Z. (1999) *Zoogeography of China*. Science Press, Beijing (in Chinese).

Zhao S., Fang J., Peng C., Tang Z. & Piao S. (2006) Patterns of fish species richness in China's lakes. *Global Ecology and Biogeography*, **15**, 386-394.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Origin information of the fish species and lakes data used for analysis in the study.

Predicting fish species richness and assemblages in Chinese lakes using Multivariate Regression Trees

Chuanbo Guo[a,b,c] , Sovan Lek[a,c], Wei Li[a], Jiashou Liu[a], Zhang Tanglin[a], Shaowen Ye*[a] ,Zhongjie Li*[a]

[a] State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China

[b] University of Chinese Academy of Sciences, Beijing 100039,China

[c] Université de Toulouse, UMR 5174 EDB, CNRS- Université de Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex 09, France

*Correspondance: Shaowen Ye & Zhongjie Li, State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China.

Email: yeshw@ihb.ac.cn; zhongjie@ihb.ac.cn

[Submitted] *Biodiversity and Conservation*

Abstract:

Predicting species richness and distribution patterns has long been the focus of theoretical and applied ecology, especially in the context of global climate change. Currently, a novel technique, Multivariate Regression Trees (MRT), have been used to predict both species richness and species distribution in order to improve the management and conservation of fish species in China. Our main objectives were (1) predicting the fish species distribution patterns and potential determinants in China; (2) identifying the species richness patterns and the drive factors; and (3) examining the predictive performance and accuracy of the novel multi-response tree model. Three accuracy measures were adopted to better evaluate the performance of the MRT: area under the curve (AUC), Cohen's Kappa, and overall accuracy. The MRT model demonstrated good or very good performance values for most of the species with high AUC, Cohen's Kappa and overall accuracy scores (median values of 0.87, 0.55 and 0.84 respectively). At the species composition level, altitude was the main determinant for fish distribution in Chinese lakes (30.43%), followed by precipitation of the driest month (10.47%), temperature annual range (3.62%) and annual mean temperature (3.15%). While at the richness level, the median AUC value of 0.5 interpret that MRT model could also better predict the fish species richness. Precipitation of driest month, maximum temperature of warmest month and lake area act as the main drivers of the fish species richness patterns. Overall, our results have provided the first comprehensive assessment of fish species composition and richness patterns on a national scale, identified the key ecological determinants, and showed both the species-energy hypothesis and species-area hypothesis to be plausible for fish species in China. Moreover, we demonstrated that MRT is a reliable and ideal community-based predictive technique for multi-species prediction.

Key words: Multi-species prediction, MRT, Fish distribution patterns, Richness patterns, China, Geographical and climatic impacts

1. Introduction

On-going global climate change has long been acknowledged to affect natural species and ecosystems (Fang et al., 2006; Buisson et al., 2010). Knowledge of species distribution and communities in ecosystems is a prerequisite for the understanding of ecosystem properties and processes as well as for effective conservation and spatial planning issues under global change (Reiss et al., 2011). Conservation and management decisions should always rely on the accurate assessment of large, or even global, scale species richness and distribution patterns, however it is too difficult to assess and monitor organisms globally. Thus species distribution models (SDM) became more important and suitable for integrating fragmental information of species and habitats (Guisan and Zimmermann, 2000). SDMs are generally developed to quantify the association between species occurrence or abundance and environmental, climatic or geographical predictors (e.g. Guisan and Zimmerman, 2000; Araújo and Guisan, 2006; Elith et al., 2007). They are usually used to explore, describe and even predict relationships between species data and environmental variables. During the last decade variations of SDMs have been substantially developed and widely applied to a variety of ecosystems, such as terrestrial ecosystems (*e.g.* Pottier et al., 2013), riverine ecosystems (*e.g.* Sharma et al., 2007) and marine ecosystems (*e.g.* Reiss et al., 2011). These models have been used to, for example, (1) explore distribution patterns or assemblages in existing habitats (e.g. Park et al., 2006; Grenouillet et al., 2011; Cheng et al., 2012), (2) evaluate the potential spread capacity of invasive species (e.g. Mika et al., 2008; Roura-Pascual et al., 2009; Brummer et al., 2013), and (3) assess biological responses to global changes (e.g. Peterson et al., 2002; Thuiller et al., 2005; Araujo et al., 2006; Buisson & Grenouillet, 2009; Tisseuil et al., 2012; Yu et al., 2013). Population modeling has increasingly emphasized prediction rather than description and explanation. Previously predictive models have been applied to the potential patterns and environmental effects of climate change on a series of taxa, including plants (e.g., Thuiller, 2004; Ohlemüller et al., 2006; Pottier et al., 2013), insects (e.g., Peterson et al., 2004), mammals (e.g., Thuiller et al. 2006a; Levinsky et al. 2007), amphibians and reptiles (e.g., Araujo et al., 2006), birds (e.g., Peterson, 2003;Virkkala et al., 2008) and fish (e.g., Bussion et al., 2008a;b; Tisseuil et al., 2012; Yu et al., 2013).

Nevertheless, most of the previous studies only focused on the species-specific model, which considers only individual species, such as generalized linear models (GLM; McCullagh & Nelder, 1983), generalized additive models (GAM; Hastie and Tibshirani, 1990), classification and regression trees (CART; Breiman et al., 1984), and random forest (RF; Breiman, 2001). As organisms are considered to co-occur as an assemblage, the species-specific models will be inadequate for multi-species data or community analysis. Guisan and Wilfried (2005) stated that SDMs currently face the challenge of reconstructing current and future patterns of communities from individual species predictions and assembly rules. It is important to fit the co-occurrence of a group of species related to environmental variables in a single process.

MRT (Multivariate Regression Trees, De'ath, 2002; Larsen and Speckman, 2004), the natural extension of CART, is an important consideration for community studies. Tree models have long been acknowledged as producing better predictions than some simple models (like LM, GLM), as tree approaches are well suited for analysis of complex ecological data (Breiman et al., 1984; De'ath and Rabricius, 2000). MRT forms clusters of sites by splitting the data in terms of environmental values, and each cluster represents a species assemblage with its environmental values being defined as its associated habitat. MRT has been widely used in ecology recently (DeVantier et al., 2006; Davidson et al., 2010; Hamann et al., 2011), yet these studies focused on the description and explanation capacity of MRT for constrained clusters definition. Until now, no attempt has been made to assess the capacity of MRT for predicting multi-species distributions, even though it is known as a predictive model for species composition when environmental variables are available (De'ath, 2002). Moreover, predictions of future scenarios of species distribution could vary considerably depending on the performance of statistical models (Araujo et al., 2005; Lawler et al., 2006; Pearson et al., 2006). If MRT is expected to be increasingly and widely used in conservation and management studies, thorough examination, including model validation and evaluation of model performance for species distribution predictions, is necessary.

Currently, in contrast with plants and birds, studies assessing the impacts of global change on fish using species distribution models are lacking. This is despite the general acceptance of species richness of fish assemblages and species distribution patterns as effective and important indicators of ecological processes as well as the state of aquatic ecosystems (Lek et al., 2005; Park et al., 2006). However, very few publications demonstrating fish species richness and distributions in Chinese lakes could be found in the literature. Importantly lakes, among the most important aquatic ecosystems on the earth, can be ideal for studying fish distribution and assemblage since they are natural bio-geographical islands. Moreover, due to the vast geographic area of China, its lakes present significantly different natural environments, climatic factors and abundance of fish species (Wang & Dou, 1998), ideal for the study of fish distributions in association with environmental and climatic variables.

Therefore, patterns at species-level (species distribution patterns) and assemblage level (species richness patterns) for fish species in Chinese lakes were modeled and predicted using MRT. In order to further assess the accuracy of MRT predictions, three different but complementary accuracy measures (overall accuracy (OA), Cohen's Kappa statistic of similarity and area under the receiver operating characteristic curve (AUC)) were implemented. Our main objectives were (1) predicting the fish species distribution patterns and potential determinants in China; (2) identifying the species richness patterns and the drive factors; and (3) examining the predictive performance and accuracy of the novel multi-response tree model.

2. Materials and methods

2.1 Study area

Lakes in China were covered with complicated natural environments and large geological variations (Wang and Dou, 1998). In the present study, data from a total of 135 lakes (21.9-48.9º N and 81.2-132.6º E, Fig.1) across mainland China with creditable fish presence-absence data were compiled. The surface areas of these lakes range from 0.0006-2933 km^2. The location and characteristics of each lake are summarized in Appendices Table S1.

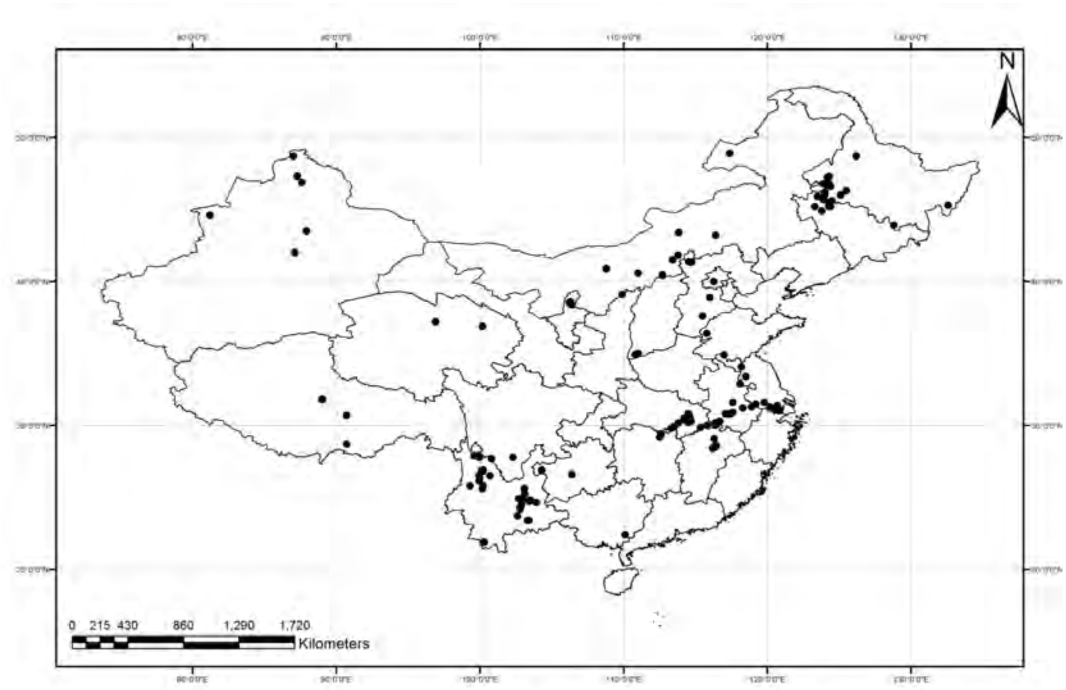


Fig.1 Locations of the study lakes in mainland China

2.2 Response variables

In accordance with the research objectives, two response variables, fish species composition and fish richness, were thus defined and predicted. Totally, 425 fish species were recorded in the 135 Chinese lakes; presence-absence data of 77 common indicator species with prevalence higher than 6% were compiled and used in the model (see Appendices Table S1 for details). Species richness index (SR), calculated for all the lakes, and was used as the richness level response variable. Presence-absence data were used because of advantages in modeling species distributions and assemblages and yielding more accurate predictions (Cawsey et al., 2002; Elith et al., 2007). To ensure the data quality and authenticity, only fish composition data derived from long-term monitoring and surveying were adopted, while data with insufficient sampling effort were omitted. For the lakes with fish data from several time periods, normally the earliest one was preferred as it was presumed that they may suffer from fewer artificial disturbances. Fish composition data were then revised according to the monograph Wu (1977; 1964) and “Fauna Sinica” to deal with possible nomenclatural changes (e.g. synonyms, newly discovered and named fish species).

2.3 Predictor variables

For the predictor variables, we used a set of climatic and geographic variables. The climatic variables were derived from the WorldClim database (Hijimans et al. 2005, available at http://www.worldclim.org), and then extracted with Arcgis 10.1 (ESRI). This database was developed from compiled monthly averages of climate as measured at weather stations from a large number of global, regional, national and local sources, mostly from 1950–2000, using the Thin Plate Smoothing Spline (TPS) algorithm that yielded climate surfaces for monthly maximum, minimum, mean temperatures and total monthly precipitation (Hutchinson 1995). In order to improve the accuracy of our analyses, we chose the highest resolution (30 arc-seconds (~1 km)). These climatic variables have been widely used in ecological modeling all over the world. However, since correlation between predictors might produce spurious results (Phillips et al. 2006), a PCA (principal component analysis) is performed before the modeling in order to eliminate the variables with high correlation. Consequently, a total of 9 environmental variables including Elevation of the site, Lake area, Annual mean temperature, Mean diurnal range (max temp–min temp), mean monthly temperature, Isothermality, Maximum temperature of warmest month, Precipitation of driest month, Precipitation of coldest quarter, are contained in the analysis (Table 1).

Table1 The description of the predictor variables used in the study (The variables with **bold** are selected to be used in the predictive model)

Variables abbreviation	**Variable type**	**Min**	**Median**	**Mean**	**Max**	**SD**
Latitude	Mean latitude	21.90	31.10	34.42	48.90	7.90
Longitude	Mean Longitude	81.20	114.40	111.60	132.60	10.47
ALT	**Elevation of the site**	2.00	144.00	911.70	4947.00	1154.56
A	Surface area of the lake	0.01	27.95	225.80	4930.00	544.82
AMT	**Annual mean temperature**	-5.00	140.00	113.70	217.00	61.45
MDR	**Mean diurnal range: mean of monthly (max temp–min temp)**	70.00	110.00	104.50	146.00	19.95
ISO	**Isothermality: (Bio2/Bio7) *100**	21.00	25.00	30.16	53.00	9.78
TS	Temperature seasonality (SD *100)	3242.00	8807.00	9305.00	15770.00	3468.65
MTWM	**Maximum temperature of warmest month**	130.00	286.00	284.40	340.00	46.71
MTCM	Minimum temperature of coldest month	-295.00	-22.00	-79.59	93.00	109.16
TAR	**Temperature annual range (Bio5–Bio6)**	212.00	331.00	364.00	561.00	103.52
MTWE	Mean temperature of wettest quarter	78.00	216.00	214.30	283.00	40.57
MTDQ	Mean temperature of driest quarter	-208.00	50.00	-2.47	162.00	104.59
MTWA	Mean temperature of warmest quarter	78.00	221.00	227.00	286.00	48.12

MTCQ	Mean temperature of coldest quarter	-213.00	43.00	-12.32	149.00	99.40
AP	Annual precipitation	82.00	919.00	806.30	1480.00	409.61
PWM	Precipitation of wettest month	18.00	186.00	168.60	305.00	63.03
PDM	**Precipitation of driest month**	0.00	9.00	15.04	44.00	14.62
PS	Precipitation seasonality(coefficient of variation)	39.00	86.00	83.35	138.00	26.64
PWEQ	Precipitation of wettest quarter	47.00	461.00	423.70	815.00	171.80
PDQ	Precipitation of driest quarter	2.00	38.00	56.64	160.00	55.12
PWAQ	Precipitation of warmest quarter	47.00	443.00	403.20	676.00	157.02
PCQ	**Precipitation of coldest quarter**	2.00	39.00	60.01	175.00	59.59

The unit for all the temperature was (°C * 10), for all the precipitation was (mm).

Min: the smallest observation**;**

Median: the median value of all the observations;

Mean: the mean value of all the observations;

Max: the largest observation;

SD: the standard deviation.

2.4 Data analysis

2.4.1 Modeling techniques

Our study was conducted at two levels: species level and assemblage level. Multivariate regression trees (MRT) are a species-driven but community-based technical approach, with regression trees as the results, where constrained clusters are always split and all species in the assemblage are explained and predicted by a series of environmental values.

MRT applies a recursive partitioning algorithm that splits objects (e.g. sampling sites) into homogenous groups according to the response, with the splits constrained by explanatory variables. The tree is grown by splitting the data a large number of times, and then it is subsequently pruned (reduction of the number of groups) via a re-sampling method called v-fold cross-validation (Breiman et al., 1984) to obtain the best predictive tree size. Each MRT model was replicated 100 times to ensure results were not obtained by chance.

The lakes were randomly subdivided into two data subsets: calibration dataset (70% of the dataset) and validation dataset (the remaining 30%). The calibration dataset was used to generate the prediction models while the validation dataset was used to evaluate the predictive performance of the model. The results of species distribution predictions from MRT modeling were then converted into binary values using one of the threshold methods, maximizing the sum of two measures: sensitivity (measures the percentage of presence correctly predicted) and specificity (measures the percentage of absence correctly predicted) (Fielding and Bell 1997).

2.4.2 Model evaluation

To better assess the performance and accuracy of the MRT models, two threshold-dependent accuracy measures (overall accuracy (OA) and Cohen’s Kappa statistic of similarity) and one threshold-independent accuracy measure (area under the receiver operating characteristic curve (AUC)) were adopted. For the prediction of species richness, only AUC was used to assess the model performance.

Overall accuracy, the probability that a site (either presence or absence) is correctly predicted, is known as the most common measure used in assessing model performance. Its application can date back to Finley (1884) who employed this measure for assessing the accuracy of tornado activity forecasts. The overall accuracy is based on binary predictions and measures the percentage of both presence and absence correctly predicted, enabling the quantification of the match between predicted and observed distributions easily with an independent data set.

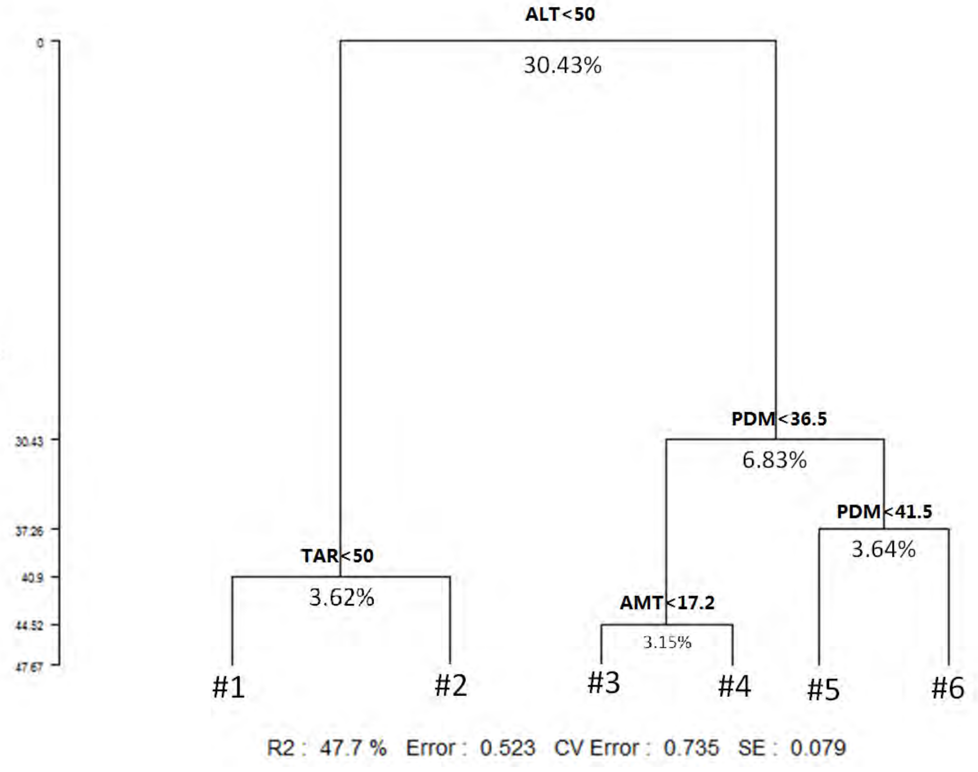
Cohen’s Kappa statistic of similarity (Cohen, 1960) is one of the most used accuracy measures in various disciplines, especially in species distribution models. Cohen’s Kappa is acknowledged to be one of the most robust evaluations and was shown to alleviate the overestimate problem of overall accuracy (Liu et al., 2011). There are classifications for the kappa value:0 - 0.2 = “Slight agreement”; 0.2 - 0.4 = "Fair agreement"; 0.4 - 0.6 = "Moderate agreement"; 0.6 - 0.8 = "Substantial agreement"; 0.8 - 1.0 = "Almost perfect agreement" (Landis and Koch, 1977).

AUC is another widely used accuracy measure in various disciplines, including ecology. However it is one of the threshold-independent accuracy measures. AUC ranges from 0 to 1, with values greater than 0.5 indicating that the model’s discrimination is no better than random, 1 implying the model discriminates perfectly (Swets 1988; Elith and Burgman, 2002).

Data analysis was conducted in R (R Core Team 2013). “MVPART wrap” (Ouellette et al. 2012) and “mvpart” (De’ath 2002) were used to extract detailed information from the MRT model, especially the output, including graphs and numerical results.

3. Results

An MRT model was constructed with six terminal nodes which can be qualified as six clusters of fish assemblages (Fig. 2a). The tree size was selected by cross-validation which showed the lowest predictive error (Fig. 2b) in the calibration dataset. The coefficient of determination (R^2) is a commonly used measure of explanatory power in linear modeling which represents the proportion of variation of the response variables explained by a model. In this model R^2=47.7%, which showed that all the variables we used in the model can explain 47.7% of the total variance. Ultimately, six fish assemblages were mainly determined by four explanation variables. Overall, altitude (ALT, 30.43%) contributed the most to fish assemblages, followed by precipitation of the driest month (PDM, 10.47%), annual temperature range (TAR, 3.62%) and annual mean temperature (AMT, 3.15%) (Fig. 2a). Assemblage 1 and 2 contained the fish species which were linked to an altitude greater than 50 m and separated by an annual temperature range lower or higher than 50℃. Fish species in assemblage 5 and 6 were related tightly at altitudes lower than 50 m, and a high precipitation during the dry season (superior than 36.5 mm). However, fish species in assemblage 3 and 4 were determined by a relatively lower altitude, and lower precipitation during the dry season (< 36.5 mm), while assemblage 3 preferred a mean annual temperature lower than assemblage 4 (Fig. 2a).



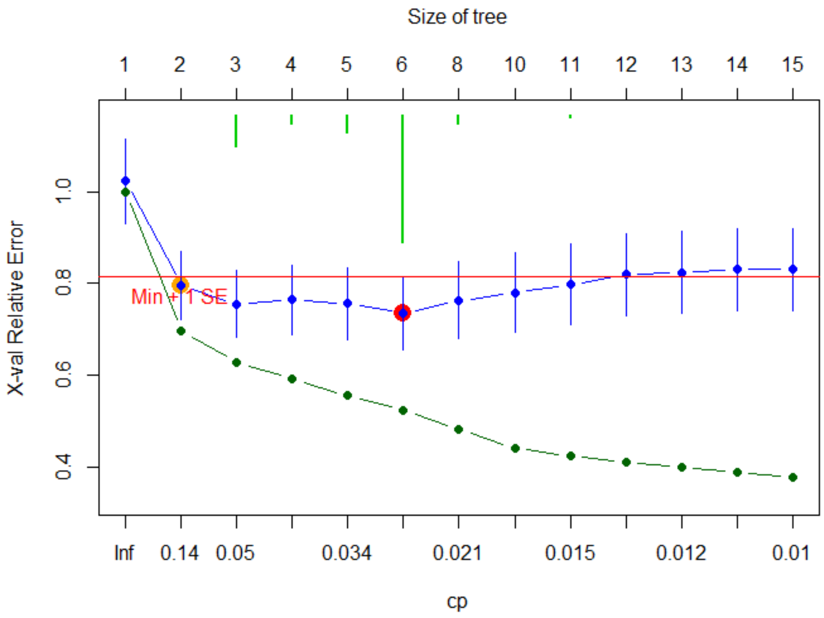


Fig.2 Multivariate response tree validated model for lake fish species distributed in China.

a. MRT tree model predicting the species distributions in Chinese lakes (see Table 1 for details of the variables)

b. Selection of the tree size by cross-validation method.

In order to evaluate the model performance, AUC (threshold-independent measure), overall accuracy (threshold-dependent) and Cohen's Kappa were used. The results showed that the MRT model could accurately predict the fish species distributions in terms of assemblage level, with AUC values ranging from 0.53 to 0.99 (with a median value of 0.87), the overall accuracy ranging from 0.61 to 0.98 (with a median value of 0.84) (Fig. 3), and the Kappa values range from 0.012 to 0.91 (with a median value of 0.55) (Fig. 4). All three accuracy measures showed significantly that most of the species were accurately predicted by the model. However three species (*Opsariichthys uncirostris bidens*, *Clarias fuscus* and *Paracheilognathus iimberbis*) were not accurately predicted, with kappa values lower than 0.2. This indicated the predictions were in 'slight agreement', while the other 74 indicator species were predicted from 'fair agreement' to 'perfect agreement' (Fig. 4). These evaluated results imply that the discrimination of the MRT model was no better than random and thus demonstrated to be robust and informative for the prediction of fish species distributions.

For species richness, a species richness index was used as the response variable. The AUC value (0.5) for the MRT model showed that the model results for species richness is no better than random and thus could be used for fish species richness prediction in Chinese lakes. Simultaneously, precipitation of driest month, maximum temperature of warmest month and lake area were identified to be the most important factors related to fish species richness in Chinese lakes. Lakes with higher precipitation of driest month (bio14⩾16.5 mm) and larger lake surface area (area ⩾ 216.1 km^2) were predicted to support more fish species, approximately 80 species per lake; while the lakes with lower precipitation of driest month (bio14 < 16.5 mm) and lower maximum temperature of warmest month (bio5 < 26 ℃) were predicted to support the fewest fish species, approximately 8 species per lake (Fig. 5).

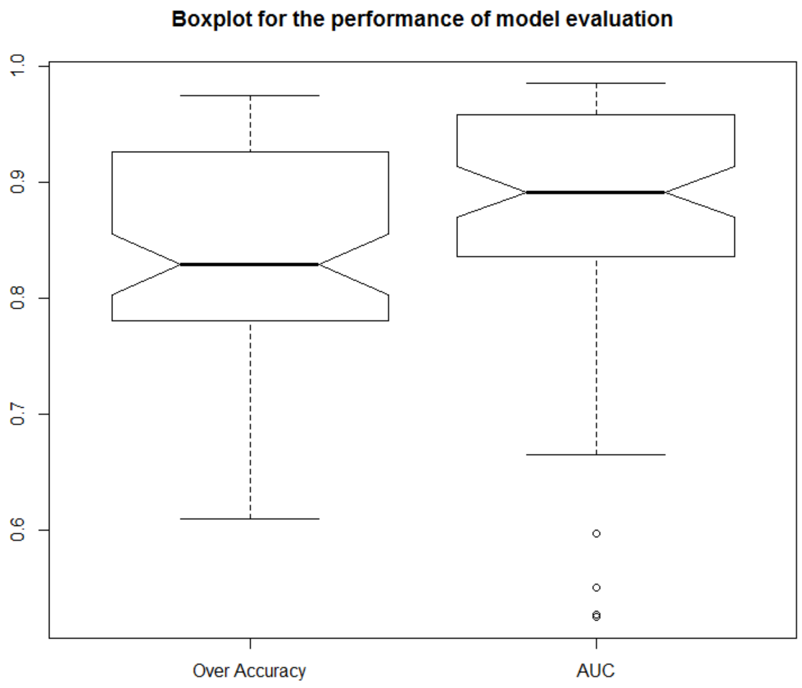


Fig.3 The evaluation results of the MRT model by Overall Accuracy and AUC measures

4. Discussion

Numerous studies have evaluated the predictive performance of different modeling techniques over the last decade. In the current study, the capacity of MRT models in predicting fish species assemblage patterns and species richness patterns based upon geographical and climatic variables was examined in Chinese lakes. Generally, we emphasize that MRT was a robust modeling technique for species prediction under global changes; ecologically, our results highlight the significance of climatic and geographical variables in determining fish assemblages and distributions in aquatic ecosystems.

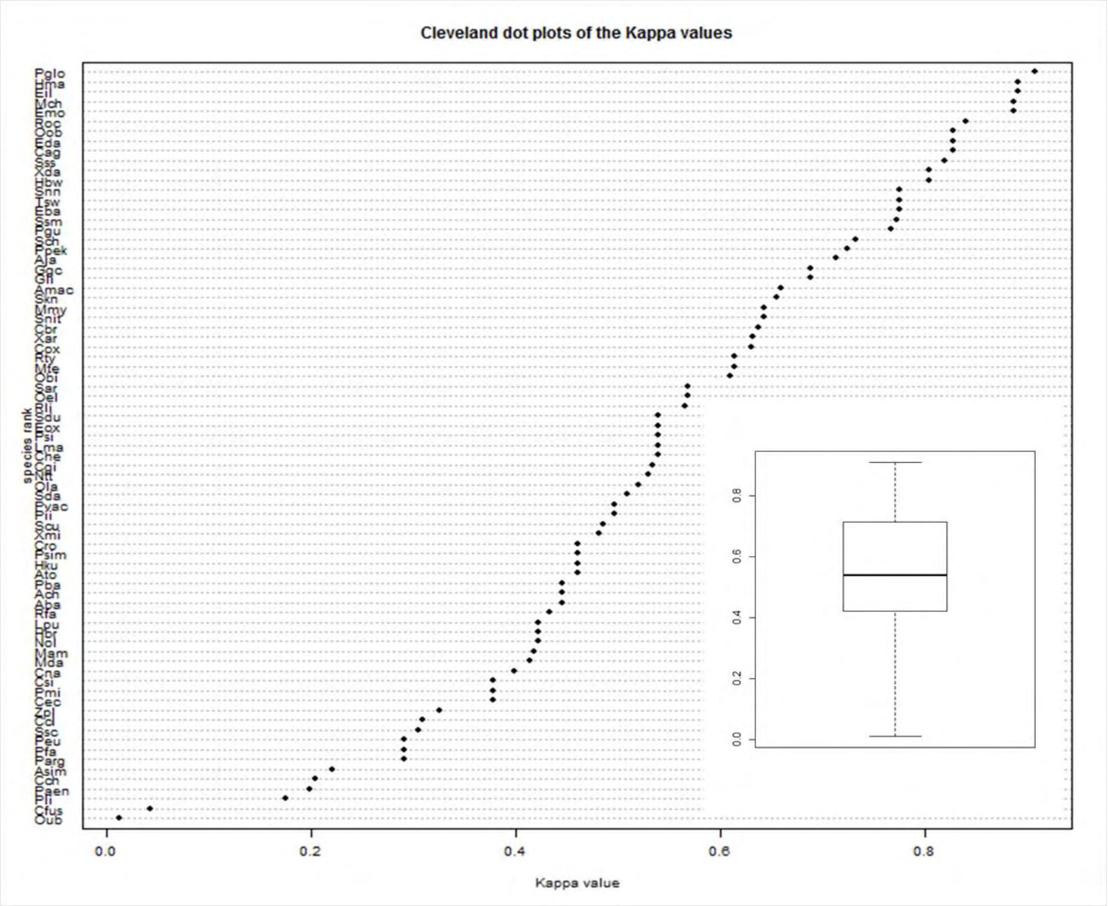


Fig.4 The species-rank kappa value for the MRT model evaluation

a. Cleveland dot plots of the kappa values

b. Boxplot of the kappa values

4.1 Model performance and technique assessment

MRT is used as the main approach for species distribution prediction. A 100 cross-validation test was employed here to strike a balance between explanatory and predictive power and to obtain a more “honest” assessment of the model, with avoid the over-fits of the multivariate regression tree analysis. Thus the minimum cross-validated error tree with six nodes was selected in this study (De’ath and Fabricus, 2000).

AUC, overall accuracy and Cohen’s Kappa were employed to better evaluate the model performance and assess the possibilities of MRT modeling for species distribution prediction. AUC is always considered as one of the most useful and popular performance measures in common ecological model studies (e.g. Manel et al. 1999, 2001; Guisan and Zimmermann 2000; Pearce and Ferrier 2000; Luck 2002; Gibson et al. 2004; Araujo et al. 2005; Luoto et al. 2006; Buisson et al, 2008a;2008b; Buisson et al. 2009; Grenouillet et al. 2011). It is considered to be the strictest performance measure by some researchers, because it is an independent threshold measure of quality (Fielding and Bell 1997; Buisson et al. 2008a). AUC scores of 0.5 indicate that a model has no discriminatory ability, while a score of 1 indicates that presence and absence are perfectly discriminated. In our study, AUC values of the indicator species varied from 0.53 to 0.99 with an average value of 0.87, indicating that almost all the species seemed to be perfectly predicted by the MRT model. The same was observed for overall accuracy, where the results varied from 0.61 to 0.98 with an average of 0.84, also indicating that all the species could be accurately predicted by the MRT model with strong discrimination.

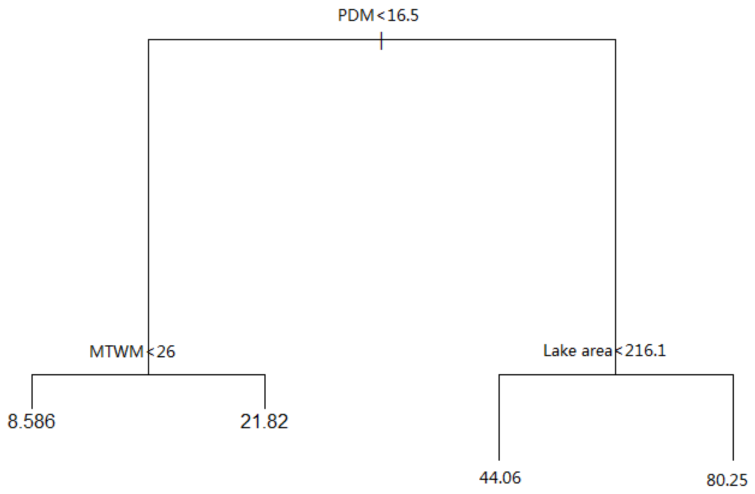


Fig.5 MRT model predicting the fish species richness in Chinese lakes. The AUC value for this model is 0.5.

Some researchers criticize that AUC and overall accuracy may overestimate model performance (Lobo et al., 2008; Peterson et al., 2008; Jimenez-Valverde, 2012). However Cohen's Kappa is considered as one of the good solutions to the overestimate problems observed with AUC and overall accuracy (Liu et al., 2011) as the Kappa index can provide a proportional accuracy for predicted presence and absence (Cohen 1960), and provide a robust evaluation of a model's performance. Kappa has already been widely used in recent ecological studies (Manel et al. 2001; Liu et al. 2005; Virkkala et al. 2005; Buisson et al. 2008a; Gevrey et al. 2009). Cohen's Kappa values in the current study varied from 0.012 to 0.91 with a mean value of 0.55. A mean Kappa value of 0.55 also indicates that most of the species were accurately predicted by the MRT model, except the three species (*Opsariichthys uncirostris bidens*, *Clarias fuscus* and *Paracheilognathus iimberbis*) which have lower kappa values between 0 and 0.2. There are many factors, such as lower prevalence, spatial autocorrelation, species attributes, environmental range size and, potentially, the model technique, which may result in the observed low accuracy of prediction for these three species.. In the current study, the results may mainly be affected by low occurrence and the environmental range of the three species. Normally for some rare species, poor performance measures would be obtained. Buisson et al. (2008a) found that rainbow trout was the rare species which obtained the low Kappa value. Actually, this phenomenon is common in some previews studies (Manel et al. 2001; Liu et al. 2005; Gevrey et al. 2009). The species' environmental range could also be one of the main factors which drive the uncertainty of the prediction of the three species in this study, as numerous studies validate that species with a smaller range can be better predicted than species with a larger environmental range (Hernandez et al., 2006; Grenouillet et al., 2011).

Overall, MRT models have been evaluated to be effective and robust enough for species distribution prediction from the ensemble evaluated measurements, even though the kappa values were marginally affected by some rare species. In conclusion, the model performance values provided optimistic estimates of the true predictive capability of MRT modeling (Araujo et al., 2005). Therefore, MRT is a robust and appropriate approach for modeling similar datasets, as it can handle complex ecological data with mixed change and high-order interactions (De'ath, 2002; Davidson et al., 2010).

Currently, although numerous studies focus on defining fish distribution patterns in conjunction with the correlated environmental factors, most are based on the species-specific models (LM, GLM, GAM, CART, RF) which typically assume that each species is independent from the others that occur at the same location (Larsen and Speckman, 2004; Guisan and Thuiller, 2005). However, in natural ecosystems, one species always coexists with many other species to form a community, within which all species are sensitive to resource competition (Wetzel, 2001). Thus, species-specific models fitted to each species would ignore the inter-specific relationships that would provide alternative and variable information about species distribution (Yin, 1990), while the prediction approach at the community scale could overcome the disadvantages of the traditional species-specific focus (Joy and De'ath, 2004; Olden et al., 2006).

4.2 Prediction and determinants of fish species assemblages in Chinese lakes

From the validated MRT model, we conclude that ALT (altitude), PDM (precipitation of driest month), TAR (temperature annual range) and AMT (annual mean temperature) were ultimately identified as the key factors determining the fish species distribution prediction in Chinese lakes, among which altitude was the most important determinant. In fact, altitude, precipitation and temperature have long been acknowledged as determinants of the distribution of organisms, such as plants (Prasad et al., 2006; Pottier et al., 2013) and riverine and marine fish and invertebrates (Bussion et al., 2008a; Cheung et al., 2009).

Altitude generally impacts distribution and life history traits of organisms globally (Amarasinghe and Welcomme 2002; Zhao et al. 2006; Yoon et al. 2011; Barradas et al. 2012; Stojkovic et al. 2013). As a complex variable, it can have direct and indirect impacts on fish distribution. Previous studies that have explored these impacts (Legendre 1993; Hawkins et al. 2003; Zhao et al. 2006), indicated that indirect impacts were even more significant than direct impacts, since so many other factors that could affect fish distribution also vary with altitude.

Without doubt, temperature is an important factor that influences fish distribution through fish metabolism (Gillooly et al., 2001), breeding (Mills & Mann, 1985), development and growth (Mann 1991; Wolter 2007) and behavior (Taniguchi et al. 1998). In our study, two kinds of temperature statistics (TAR and AMT) were defined as important in the determination of fish distribution, which is consistent with Crisp (1996) and Mann (1996) who considered these two factors as of high importance for the biological requirements of freshwater fish. The influence from TAR in the present study may be partly due to the vast variability of the temperature (21.2 - 56.1℃), as fish are poikilothermal animals and therefore highly sensitive to changes in temperature. Some European freshwater fishes (Buisson et al., 2008a; Pont et al., 2005; Crisp, 1996; Mann, 1996) and American species (Rathert et al., 1999) were also demonstrated to be great impacted by the TAR. Although each fish species has a thermal preference characterized by a tolerance range of temperatures, they are flexible and can adapt to a certain extent with physiological adjustments (Johnson &

Kelsch, 1998) or behavioral thermoregulation (Heggenes et al., 1993). However, thermal ranges exceeding the normal range will impact fish populations and distributions (Gale et al., 2013; Gislason et al., 2010). Extremely low temperatures may affect metabolism, breeding, growth, behavior, and thereby fish distributions (Ruibidge et al., 2011; Yu et al., 2013; Aguilar-Kirigin & Naya, 2013).

Apart from altitude and temperature, precipitation has been considered as one of the most important climatic factors in numerous recent studies (Zhao et al., 2006; Buisson et al., 2008a;b; Buisson and Grenouillet, 2009). Precipitation may impact stream flows and hydrological conditions, while hydrology is even more complex and concerns diverse factors such as meteorology, geomorphology, geology or geography, which could all affect fish assemblages and distribution. The PDM is also considered as having a large impact on niche shifts of freshwater species (Lauzeral et al., 2011), and thereby affecting fish distribution. Indeed, the PDM can modify the hydrological conditions of lakes directly, such as water content, water depth and transparency, nutrient supplements (Wetzel, 2001), which in turn affect fish populations and distributions (Brazner and Beals, 1997; Petry et al., 2003; Mello et al., 2009; Cheng et al., 2012; Kang et al., 2013).

The set of the eight variables used in this study was relatively successful in predicting and explaining the fish assemblages using the MRT model. However, a number of factors have been identified as potentially affecting fish distribution and diversity globally, such as land-cover, geographic variables, topological variables, biological invasion, artificial impacts and net primary production (Park et al., 2006; Buisson et al., 2008a,b; Gevrey et al., 2009; Maloney et al., 2013; Kang et al., 2013). Further research therefore requires the inclusion of such factors in order to obtain more powerful explanations. More importantly, as we live in a changing world, predictions of species distributions based on possible future scenarios are more desirable than the explanation of the current situation. Thus, more efforts should be made to improve the understanding of the potential distribution and impacts of fish species associated with the global climate change.

4.3 Prediction and determinants of fish species richness in Chinese lakes

Species richness is one of the most important biodiversity indicators in ecology. At this level, our research revealed that fish species richness was highly related to precipitation of driest month (PDM), maximum temperature of warmest month (MTWM) and lake area (A). As seen previously, fish species distributions are highly related to precipitation and temperature, it therefore follows that species richness should also be affected by precipitation and temperature. Results also showed that higher fish species richness is predicted in lakes with sufficient precipitation and a larger surface area. This is in agreement with the actual situation in China where lakes distributed along the middle reach of the Yangtze River, such as Dongting lake, Liangzi lake, Hongze lake, Poyang lake, have a higher fish richness as they are exposed to an appropriate temperature, abundant rainfall and are lakes with relatively larger areas (Cui and Li, 2005). However, lakes distributed in parts of Qinghai-Tibet, Meng-Xing and north-east China have relatively lower fish richness due to the lower precipitation and lower temperature (Zhao et al., 2006). In fact, several hypotheses have been presented on species richness patterns, among which the species-area hypothesis and the species-energy hypothesis (Wright, 1983) are the most popular. In this study, we found that PDM, MTWM and lake area best explained the variability of fish species richness patterns in China. These results therefore support the plausibility of both hypotheses for the fish of Chinese lakes. Our results support the species-energy hypothesis, which claims that environments with higher energy availability could support more species, as well as supporting the hypothesis that large surface areas could support more fish species, although in this study lake area was not the highest contributer to the prediction of species richness. The lake area effects may be overshadowed by the large variations of the climatic variables in our study. Our findings were generally consistent with some studies in terrestrial ecosystems (Wright et al., 1999; Hawkins et al., 2003a), and freshwater ecosystems (Zhao et al., 2006). Some other studies, with a smaller scale focus,, also argued that fish species richness was highly related with water depth (Cheng et al., 2012), habitat (Petry et al., 2003) and other organisms (Xie et al., 2001) in lakes.

The results of the current study may provide a basis for future research using MRT as well as contributing to conservation of fish biodiversity under global changes.

In conclusion, our research focuses on the understanding of fish species richness and distributions patterns as well as the potential driving factors behind these parameters in Chinese lakes. The recursive partitioning and regression tree models MRT were used and thus be examined. Simultaneously, we have drawn regarding the relative roles of the environmental and climatic variables in driving the lake fish species distribution and richness patterns. Further study of this approach using more accurate predictive models should be encouraged to advance our understanding of the profound influence of global change on species distribution under future scenarios.

Acknowledgements

Many thanks should be given to Radika Michniewicz for her kind help with the English corrections. This research was supported by Grant Nos. 200903048 and 20130305 from the “Special Fund for Agro-scientific Research in the Public Interest” of China, Nos. 30830025 and 30900182 from the National Natural Science Foundation of China. We are grateful to the China Scholarship Council (CSC) for their financial support of Chuanbo Guo’s study abroad.

Appendices

Table S1 The origin information of fish data and lakes used in this study

References

Aguilar-Kirigin ÁJ, Naya DE. Latitudinal patterns in phenotypic plasticity: the case of seasonal flexibility in lizards' fat body size. Oecologia 2013; 1-8.

Amarasinghe US, Welcomme RL. An analysis of fish species richness in natural lakes. Environ Biol Fish 2002; 65(3), 327-339.

Araújo MB, Pearson RG, Thuiller W, Erhard M. Validation of species–climate impact models under climate change. Glob Change Biol 2005; 11: 1504-1513.

Barradas JRS, Silva L., Harvey BC, Fontoura NF. Estimating migratory fish distribution from altitude and basin area: a case study in a large Neotropical river. Freshwater Biol 2012; 57: 2297-2305.

Brazner J, Beals E. Patterns in fish assemblages from coastal wetland and beach habitats in Green Bay, Lake Michigan: a multivariate analysis of abiotic and biotic forcing factors. Can J Fish Aquat SCI 1997; 54: 1743－1761.

Breiman L. Random forests. Mach Learn 2001*;* **45**: 5-32.

Breiman L, Friedman JH, Olshen RA, Stone CG. Classification and regression trees. Wadsworth International Group, Belmont, California, USA; 1984

Brummer TJ, Maxwell BD, Higgs MD, Rew LJ. Implementing and interpreting local－scale invasive species distribution models. Divers Distrib 2013; doi: 10.1111/ddi.12043.

Buisson L, Grenouillet G. Contrasted impacts of climate change on stream fish assemblages along an environmental gradient. Divers Distrib 2009*;*15: 613–626.

Buisson L, Grenouillet G, Casajus N, Lek S. Predicting the potential impacts of climate change on stream fish assemblages. Am Fish S S 2010; 73 : 327-346.

Buisson L, Thuiller W, Lek S, Lim P, Grenouillet G. Climate change hastens the turnover of stream fish assemblages. Glob Change Biol 2008a*;*14: 2232–2248.

Buisson L, Blanc L, Grenouillet G. Modelling stream fish species distribution in a river network: the relative effects of temperature versus physical factors. Ecol Freshw Fish 2008b;17:244－257.

Cawsey EM, Austin MP, Baker BL. Regional vegetation mapping in Australia: a case study in the practical use of statistical modeling. Biodivers Conserv 2002;

11:2239–2274.

Cheng L, Lek S, Lek-Ang S, Li Z. Predicting fish assemblages and diversity in shallow lakes in the Yangtze River basin. Limnologica 2012;42: 127-136.

Cheung WW, Lam VW, Sarmiento JL, Kearney K, Watson R, Pauly D. Projecting global marine biodiversity impacts under climate change scenarios. Fish Fish 2009: 10: 235-251.

Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960: 20:37-46.

Crisp DT. Environmental requirements of common riverine European salmonid fish species in freshwater with particular reference to physical and chemical aspects. Hydrobiologia 1996: 323:201－221.

Davidson TA, Sayer CD, Langdon PG, Burgess A, Jackson M. Inferring past zooplanktivorous fish and macrophyte density in a shallow lake: application of a new regression tree model. Freshw Biol 2010;55:584-599.

De'ath G. Multivariate regression trees: a new technique for modeling species-environment relationships. Ecology 2002: 83: 1105-1117.

De'ath G, Fabricius KE. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 2000: 81: 3178-3192.

DeVantier LM, De’Ath G, Turak E, Done TJ, Fabricius KE. Species richness and community structure of reef-building corals on the nearshore Great Barrier Reef. Coral Reefs 2006:25: 329-340.

Elith J, Burgman MA. Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia; 2002

Elith J, Leathwick J. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. Divers Distrib 2007:13:265－275.

Fang JY, Wang ZH, Zhao SQ, Li YK, Tang ZY, Yu D, Ni LY, Liu HZ, Xie P, Da LJ, Li ZQ, Zheng CY. Biodiversity changes in the lakes of the Central Yangtze. Front Ecol Environ 2006: 4: 369-377.

Fielding AH, Bell JF. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Cconserv 1997: 24: 38-49.

Finley JP. Tornado prediction. Amer Meteor J; 1884.

Gale MK, Hinch SG, Donaldson MR. The role of temperature in the capture and release of fish. Fish Fish 2013*;* 14: 1-33.

Gevery M, Sans-Piche F, Grenouillet G, Tudesque L,Lek S. Modeling the impact of landscape types on the distribution of stream fish species. Can J Fish Aquat Sci 2009; 66: 484-495.

Gibson LA, Wilson BA, Cahill DM, Hill J. Spatial prediction of rufous bristlebird habitat in a coastal heathland: a GIS‐based approach. J Appl Ecol 2004; 41: 213-223.

Gillooly JF, Brown JH, West GB, Savage Van M, Charnov EL. Effects of size and temperature on metabolic rate. *Science* 2001;293: 2248–2251.

Gislason H, Daan N, Rice JC, Pope JG. Size, growth, temperature and the natural mortality of marine fish. Fish Fish, 2010;11: 149–158.

Grenouillet G, Buisson L, Casajus N, Lek S. Ensemble modelling of species distribution: the effects of geographical and environmental ranges. Ecography 2011; 34: 9-17.

Guisan A ,Zimmerman NE. Predictive habitat distribution models in ecology. Ecol Model 2000; 135:147-186.

Guisan A, Thuiller W. Predicting species distribution: offering more than simple habitat models. Ecol Lett 2005; 8: 993-1009.

Hamann A, Gylander T, Chen PY. Developing seed zones and transfer guidelines with multivariate regression trees. Tree Genet Genomos 2011;7: 399-408.

Hastie TJ, Tibshirani RJ.*Generalized additive models*. Chapman and Hall, London, UK; 1990.

Hawkins BA, Field R, Cornell HV, Currie DJ, Guégan JF, Kaufman DM, ... Turner JR. Energy, water, and broad-scale geographic patterns of species richness. Ecology 2003; 84:3105-3117.

Heggenes J, Krog OMW, Lindas OR, Dokk JG & Bremnes T. Homeostatic behavioural responses in a changing environment: brown trout (*Salmo trutta*) become nocturnal during winter. J Anim Ecol 1993:62:295-308.

Heikkinen RK, Luoto M, Araujo MB, Virkkala R, Thuiller W, Sykes MT (2006) Methods and uncertainties in bioclimatic envelope modelling under climate change. Progress in Physical Geography, 30, 751–777.

Hijmans RJ, Cameron SE, Parra JL, Jones P.G & Jarvis A. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 2005;25:1965–1978.

Hutchinson MF. Interpolating mean rainfall using thin plate smoothing splines. Int J Geogr Inf syst 1995; 9:385-403.

Jiménez - Valverde A. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. Glob Ecol Biogeogr 2012; 21: 498-507.

Johnson JA, & Kelsch SW. Effects of evolutionary thermal environment on temperature-preference relationships in fishes. Environ Biol Fish 1998;53: 447-458.

Joy MK, & Death RG. Predictive modelling and spatial mapping of freshwater fish and decapod assemblages using GIS and neural networks. Freshw Biol 2004; 49: 1036-1052.

Kang B, Deng J, Wu Y, Chen L, Zhang J, Qiu H, Lu Y & He D. Mapping China's freshwater fishes: diversity and biogeography. *Fish Fish;*2013, doi: 10.1111/faf.12011

Landis JR, & Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 ;159-174.

Larsen DR, & Speckman PL. Multivariate regression trees for analysis of abundance data. Biometrics 2004; 60: 543-549.

Lauzeral C, Leprieur F, Beauchard O, Duron Q, Oberdorff T,, Brosse S. Identifying climatic niche shifts using coarse-grained occurrence data: a test with non‐native freshwater fish. Glob Ecol Biogeogr 2011;20: 407-414.

Lawler JJ, White D, Neilson RP, Blaustein AR. Predicting climate‐induced range shifts: model differences and model reliability. Glob Change Biol 2006; 12: 1568-1584.

Legendre P. Spatial autocorrelation: trouble or new paradigm? Ecology 1993; 74: 1659-1673.

Levinsky I, Skov F, Svenning JC, Rahbek C. Potential impacts of climate change on the distributions and diversity patterns of European mammals. Biodivers Conserv 2007; 16: 3803-3816.

Liu C, Berry PM, Dawson TP, Pearson RG. Selecting thresholds of occurrence in the prediction of species distributions. Ecography 2005; 28: 385-393.

Liu C, White M, Newell G. Measuring and comparing the accuracy of species distribution models with presence–absence data. Ecography 2011;34: 232-243.

Lobo JM, Jiménez‐Valverde A, Real R. AUC: a misleading measure of the performance of predictive distribution models. Glob Ecol Biogeogr 2008:17: 145-151.

Luck GW. The habitat requirements of the rufous treecreeper (Climacteris rufa). 1. Preferential habitat use demonstrated at multiple spatial scales. Biol Conserv 2002;105: 383-394.

Luoto M, Heikkinen RK, Pöyr J, Saarinen K. Determinants of the biogeographical distribution of butterflies in boreal regions. J Biogeogr 2006: 33:1764-1778.

Maloney KO, Weller DE, Michaelson DE, Ciccotto PJ. Species Distribution Models of Freshwater Stream Fishes in Maryland and Their Implications for Management. Environ Model Assess 2013;18:1-12.

Manel S, Dias JM, Ormerod SJ. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. Ecol Model 1999;120: 337-347.

Manel S, Williams HC, Ormerod SJ. Evaluating presence–absence models in ecology: the need to account for prevalence. J Appl Ecol 2001;38: 921-931.

Mann RHK. Environmental requirements of European non-salmonid fish in rivers. Hydrobiologia. 1996; 323: 223-235.

McCullagh P, Nelder JA. *Generalized linear models*. Chapman and Hall, London, UK;1983

Mello TD, Meerhoff M, PEKCAN‐HEKIM, ZEYNEP, Jeppesen E. Substantial differences in littoral fish community structure and dynamics in subtropical and temperate shallow lakes. Freshw Biol 2009; 54:1202-1215.

Mika AM, Weiss RM, Olfert O, Hallett RH, Newman JA. Will climate change be beneficial or detrimental to the invasive Swede midge in North America? Contrasting predictions using climate projections from different general circulation models. Glob Change Biol 2008; 14: 1721-1733.

Mills CA, Mann RHK.Environmentally-induced fluctuations in year-class strength and their implications for management. J Fish Biol 1985; 27: 209 – 226.

OhlemÜller R, Gritti ES, Sykes MT, Thomas CD. Quantifying components of risk for European woody species under climate change. Glob Change Biol 2006; 12: 1788-1799.

Olden JD, Rooney TP. On defining and quantifying biotic homogenization. Global Ecol Biogeogr 2006; 15: 113-120.

Ouellette MH, Legendre P, Borcard D. Cascade multivariate regression tree: a novel approach for modelling nested explanatory sets. Method Ecol Evol 2012; 3: 234-244.

Park YS, Grenouillet G, Esperance B, Lek S. Stream fish assemblages and basin land cover in a river network. Sci Total Environ 2006; 365: 140-153.

Pearce J, Ferrier S. Evaluating the predictive performance of habitat models developed using logistic regression. Ecol model 2000; 133: 225-245.

Pearson RG, Dawson TP. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Glob Ecol Biogeogr 2003; 12: 361-371.

Pearson RG, Thuiller W, Araújo MB, Martinez‐Meyer E, Brotons L, McClean C, ... Lees DC. Model-based uncertainty in species range prediction. J Biogeogr 2006; 33: 1704-1711.

Peterson AT. Predicting the geography of species' invasions via ecological niche modeling. The quarterly review of Biol 2003; 78: 419-433.

Peterson AT, Papeş M, Soberón J. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecol Model 2008; 213: 63-72.

Peterson BJ, Holmes RM, McClelland JW, Vorosmarty CJ, Lammers RB, Shiklomanov AI., ... Rahmstorf S. Increasing river discharge to the Arctic Ocean. *Science* 2002;298 : 2171-2173.

Petry P, Bayley P, Markle D. Relationships between fish assemblages, macrophytes and environmental gradients in the Amazon River floodplain. J Fish Biol 2003; 63: 547-579.

Phillips SJ, Anderson RP, Schapire RE. Maximum entropy modeling of species geographic distributions. Ecol model 2006; 190: 231-259.

Pont D, Hugueny B, Oberdorff T. Modelling habitat requirement of European fishes: do species have similar responses to local and regional constraints? Can J Fish Aquat Sci 2005; 62 :163-173.

Pottier J, Dubuis A, Pellissier L, Maiorano L, Rossier L, Randin CF, ... Guisan A. The accuracy of plant assemblage prediction from species distribution models varies along environmental gradients. Global Ecol Biogeogr 2013; 22:52-63.

Prasad AK, Singh RP, Kafatos M. Influence of coal based thermal power plants on aerosol optical properties in the Indo‐Gangetic basin. Geophys Res Lett 2006: 33

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rathert D, White D, Sifneos JC, Hughes RM. Environmental correlates of species richness for native freshwater fish in Oregon, USA. J Biogeogr 1999;26: 257–273.

Reiss H, Cunze S, Koenig K, Neumann H, Kroencke I. Species distribution modelling of marine benthos: a North Sea case study. Mar Ecol Prog Ser 2011; 442: 71-86.

Roura-Pascual N, Brotons L, Peterson AT, Thuiller W. Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. Biol Invasions 2009 ;11: 1017-1031.

Rubidge EM, Monanan WB, Parra JL, Cameron SE, Brashares JS. The role of climate, habitat, and species cooccurrence as drivers of change in small mammal distributions over the past century. Glob Change Biol 2011;17: 696-708.

Sharma S, Jackson DA, Minns CK, Shuter BJ. Will northern fish populations be in hot water because of climate change? Glob Change Biol 2007;13:2052–2064.

Stojkovic M, Simic V, Milosevic D, Mancev D, Penczak T. Visualization of fish community distribution patterns using the self-organizing map: A case study of the Great Morava River system (Serbia). Ecol Model 2013; 248:20-29.

Swets JA Measuring the accuracy of diagnostic systems. Science 1988;240: 1285-1293.

Taniguchi Y, Rahel FJ, Novingen DC, Gerow KG. Temperature mediation of competitive interactions among three fish species that replace each other along longitudinal stream gradients. Can J Fish Aquat Sci 1998;:1894-1901.

Thuiller W. Patterns and uncertainties of species' range shifts under climate change. Glob Change Biol 2004;10: 2020-2027.

Thuiller W, Broennimann O, Hughes G, ALKEMADE JRM, MIDGLEY GF, Corsi F. Vulnerability of African mammals to anthropogenic climate change under conservative land transformation assumptions. Glob Change Biol 2006; 12: 424-440.

Thuiller W., Lavorel, S., Araújo, M. B., Sykes, M. T., & Prentice, I. C. Climate change threats to plant diversity in Europe. P Natl Acad Sci USA 2005; 102 :8245-8250.

Tisseuil C, Leprieur F, Grenouillet G, Vrac M, Lek S. Projected impacts of climate change on spatio‐temporal patterns of freshwater fish beta diversity: a deconstructing approach. Global Ecol Biogeogr 2012; 21: 1213-1222.

Virkkala R, Heikkinen RK, Leikola N, Luoto M. Projected large-scale range reductions of northern-boreal land bird species due to climate change. Biol Conserv 2008;141: 1343-1353.

Virkkala R, Luoto M, Heikkinen RK, Leikola N. Distribution patterns of boreal marshland birds: modelling the relationships to land cover and climate. J Biogeogr 2005;32: 1957-1970.

Wang SM, Dou HS. Lakes of China. *Science Press*, Beijing;1998

Wetzel RG. Limnology: lake and river ecosystems. *Academic press*. 2001

Wolter C. Temperature influence on the fish assemblage structure in a large lowland river, the lower Oder River, Germany. Ecol freshwater fish 2007; 16: 493-503.

Wu XW. The Cyprinoid Fishes of China, Vol. 1. People’s Press, Beijing;1964 (in Chinese).

Wu XW. The Cyprinoid Fishes of China, Vol. 2. People’s Press, Beijing;1977 (in Chinese).

Yin M. Ecology of fish [M].China agriculture Press, Beijing;1982 (In Chinese)

Xie S., Cui Y. & Li Z. 2001. Small fish communities in two regions of the Liangzi Lake, China, with or without submersed macrophytes. J App Ich 17: 89-92.

Yoon JD, Kim JH, Byeon M., Yang HJ, Park JY, Shim JH., ... Jang MH. Distribution patterns of fish communities with respect to environmental gradients in Korean streams. Ann Limnol-Int J Lim2011; 47: 63-71.

Yu D, Chen M, Zhou Z, Eric R, Tang Q, Liu H. Global climate change will severely decrease potential distribution of the East Asian coldwater fish Rhynchocypris oxycephalus (*Actinopterygii*, *Cyprinidae*). Hydrobiologia 2013;700:23-32.

Zhao S, Fang J, Peng C, Tang Z, Piao S. Patterns of fish species richness in China's lakes. Glob Ecol Biogeogr 2006; 15: 386-394.

Uncertainty in ensemble modelling of large-scale species distribution: effects from the species characteristics and model techniques

Chuanbo Guo[a,b,c], Sovan Lek[a,b] , Shaowen Ye[b], Wei Li[b], Jiashou Liu[b], Zhongjie Li[b*]

a. Université de Toulouse, UMR 5174 EDB, CNRS, ENFA, Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex 09, France

b. State Key laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China

c. University of the Chinese Academy of Sciences, Beijing 100039, China

* Corresponding author. Tel.: +86 27 68 78 00 63; fax: +86 27 68 78 00 63.

E-mail address: zhongjie@ihb.ac.cn (Z. Li)

Submitted to [*Ecological Modelling]*

Abstract:

Species distribution models (SDM) have been routinely used for the purpose of species conservation and biodiversity management, especially in the context of global climate change. However, there is little knowledge about the uncertainty source on the SDM for the predictions in aquatic ecosystems, especially in the large-scale research. Therefore, we contribute to the first perspective on the uncertainties of SDMs in predicting fish species distribution in lake ecosystems. Totally, 92 fish species were predicted with climatic and geographical variables respectively using nine widely implemented species distribution models. Generally, we focused on explore the potential impacts from two main kinds of uncertainty sources: species characteristics (contained species prevalence, altitude range, temperature range and precipitation range) and model technique (calibration technique and evaluation technique). Finally, our results highlight that predictions from single SDM were so variety and unreliable for all species while ensemble approaches could yield more accurate predictions; we also found that there was no significant influence on the model outcomes from the evaluation measures; we emphasized that species characteristics as species prevalence, altitude range size and precipitation range size would strongly affect the outcomes of SDMs, but temperature range size didn’t show a significantly influence; our findings finally verified the hypothesis that species distributed with a smaller range size could be more accurately predicted than species with large range size to be plausible in aquatic ecosystems. Our research would provide promising insights into the predicting of fish species in aquatic ecosystems under the impacts of global climate change, especially for the conservation of endemic fish species in China which we inferred could be better predicted. Moreover, our results improved the understanding of uncertainties from species characteristics and model techniques in species distribution model.

Keywords: Ensemble models; Uncertainties; Species characteristic; model technique; China; Fish species distribution and assemblage

Introduction

Species distribution model (SDM), generally developed to quantify the association between species' occurrence or abundance and environmental, climatic or geographical predictors (e.g. Guisan and Zimmerman, 2000; Araújo and Guisan, 2006; Elith et al., 2007), have been widely implemented in both basic and applied ecology recently, especially for species conservation and biodiversity management. Over the last decade, numerous of SDMs as Generalized Linear Models (GLMs, McCullagh and Nelder, 1989), Generalized Additive Models (GAMs, Hastie and Tibshirani, 1990), Classification and Regression Tree (CART, Breiman et al., 1984), Random Forest (RF, Breiman, 2001), and even some multi-species model as Multivariate Adaptive Regression Splines (MARS, Friedman, 1991), Artificial Neural Networks (ANN, Ripley, 1996) and Multivariate Regression Trees (MRT, De'ath, 2002) have been documented and applied routinely for (1) predicting distribution patterns or assemblages from the current habitat status (e.g. Park et al., 2006; Buisson et al., 2008; Grenouillet et al., 2011), (2) evaluating the potential spreading capacity of invasive species (e.g. Roura-Pascual et al., 2009; Brummer et al., 2013), and (3) assessing biological responses and occurrences to global changes (e.g. Thuiller et al., 2005; Araujo et al., 2006; Buisson and Grenouillet, 2009; Tisseuil et al., 2012; Yu et al., 2013). However, variability and uncertainty about the outcomes of different statistical models, has been rarely considered and assessed.

To date, along with the widely application of SDMs, increasingly publications began to focus on the accuracy and uncertainty of species prediction, results demonstrated that predictions from SDMs were not equally reliable for all species and thus the best performing models are not always the same for different species (Luoto et al., 2005; Segurado and Araujo, 2004; Barbet-Massin et al., 2009). Thus the outcomes and ecological predictions derived from alternative single models could be so variety as to affect the decision making (Guisan et al., 2007; Araujo and New, 2007; Buisson et al., 2010), while quantify the contribution from different sources of uncertainty will help to reduce the variance of ecological predictions and decide where to take into account in the future research to reduce variability and improve the reliability in projections (Buisson et al., 2010). Due to the aforementioned issues, former studies have also managed to explore the uncertainty sources of the species distribution modelling. Up until now, model techniques, data characteristics, species prevalence, latitudinal range, spatial autocorrelation, rarity and species environmental ranges have all considered to be the potential uncertainty source which could affect the model performance (Manel et al., 2001; Segurado and Araujo, 2004; Boone and Krohn, 1999; Karl et al., 2000; Marmion et al., 2009b). However, very few researches quantify both of the uncertainties from species characteristic and model techniques synthetically, especially in aquatic ecosystems (But see Grenouillet et al., 2011), In addition, former researches only conducted in a small geographical scale, which would be constrained by the country border and thus hide the effects of geographical and environmental ranges (Grenouillet et al., 2011). While it is worth noting that the factors determine species distributions vary according to the scale of analysis (Whittaker et al., 2001). Nowadays, large-scale study of species distribution prediction in aquatic ecosystems could rarely be found.

Over the last two decades, SDMs have been widely used across terrestrial ecosystem, stream ecosystems, marine systems, while lakes remain poorly studied despite the fact that they are among the most important aquatic ecosystems on earth and can be ideal fields for studying fish distribution and assemblage since they are natural bio-geographical islands (Wang and Dou, 1998; Zhao et al., 2006). Moreover, due to the large potential uncertainty sources existed in modeling the species distribution in aquatic ecosystems, there is nearly no evidence could be found for fish species prediction using SDMs in lake ecosystems in the world. Therefore, assessment of the uncertainty source and accuracy on the species distribution models in predicting fish distribution and diversity in lakes should be urgently addressed before the projections were used for decision making in the context of global change.

Consequently, in this context, we designed our study which takes into account both of the uncertainties from species characteristics (species prevalence, altitude range, temperature range and precipitation range) and model techniques (9 SDMs × 3 evaluation methods) in a large geographical scale in aquatic ecosystems. Finally, a distribution of 92 fish species in 135 Chinese lakes were modeled with nine widely used species distribution models (GLM, RF, CTA, GAM, MARS, SRE, ANN, GBM and MDA), the performance were evaluated with Area under the receiver operating characteristic curve (AUC ; Fielding and Bell, 1997), Cohen’s Kappa (Cohen, 1960) and True skill statistic (TSS ; Bohning et al., 2008). Generally, our main objectives were 1) to assess the accuracy and uncertainties of nine widespread species distribution models in the ensemble framework; 2) to quantify the relative contributions of different uncertainty sources on the ecological predictions; 3) to test the hypothesis that species with specialized range size could be more accurately predicted in aquatic ecosystems. We contributed to the better understanding of the uncertainties on species distribution models and make good basement for the future projection of fish species in the climate change context.

Materials and methods

Study area

The mainland China spanned over a large geographic area, and the lakes distributed across the country with significant different geographical and climatic conditions. Therefore, Chinese lakes, which occupy significantly different natural environments, climatic factors and fish biota due to the vast area, make them more than suitable for the case study of fish diversity and distributions in relation to environment and climatic variations in a large scale (Wang and Dou, 1998). To be comprehensively, total of 135 lakes with the locations spanned over 21.9□ - 48.9□ N and 81.2□ - 132.6□ E were extracted in the present study, the surface area of each lake ranged from 0.0006 km^2-2933 km^2 , the location and limnological characteristics were precisely summarized in Table S1. Most of all, these lakes were selected due to the available of high quality fish data.

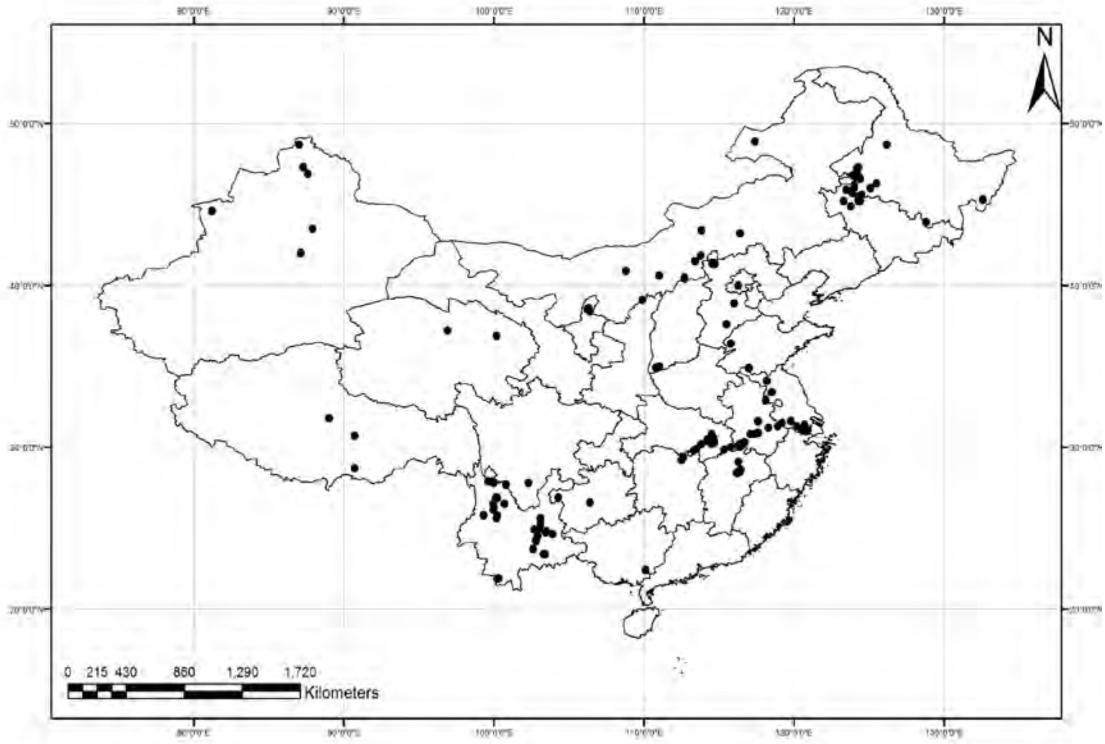


Fig.1 Map of study lakes distributed in mainland China

Fish data

In the current study, presence-absence data of fish in the 135 lakes were collected from published fish surveys since 1950s (see Appendix Table S1). However, in order to ensure the data quality and authenticity, fish assemblages' data should be derived from long-term monitoring and surveying, the data without enough sampling efforts were omitted. For the lakes which have fish data surveyed in several time periods, normally the earliest dataset was preferred as they may suffer from aritificial disturbances. Fish assemblages' data were also sifted seriously and then corrected according to the "*Fauna Sinica*" and Wu (1977; 1980) to integrate the species with confused names. Totally, 425 fish species and subspecies were contained in all the study lakes with presence-absence data compiled from published works, cause these data were proved to be more ideal for modelling species distributions and assemblages (Elith et al., 2007; Cawsey et al., 2002). However, to be efficiently, only fish species with occurrence than 10 times in all the sites were used in SDMs, therefore 92 fish species were ultimately predicted.

Climatic and environmental variables

For the explanatory variables, we used a set of climatic variables and lake geographic variables. The climatic variables derived directly from the WorldClim database (Hijimans et al., 2005, available at http://www.worldclim.org), and then extracted use Arcgis 10.1 (ESRI). This database was developed from compiled monthly averages of climate as measured at weather stations from a large number of global, regional, national and local sources, mostly from the 1950–2000 periods, using the Thin Plate Smoothing Spline (TPS) algorithm that yielded climate surfaces for monthly maximum, minimum, mean temperatures and total monthly precipitation (Hutchinson, 1995). In order to improve the accuracy of our analyses, we choose the highest resolution (30 arc-seconds (~1 km)); at last, 19 bioclim-variables as well as the altitude of the lake were extracted from the dataset. Moreover, lake surface area was also contained as an explanatory factor in the data set (Table 1). These climatic variables have been widely used in ecological modeling all over the world. However, due to some of them may provide similar environmental information, a PCA (principal component analysis) is performed to eliminate the high correlate and redundant variables. Since correlation between predictors might produce spurious results (Phillips et al., 2006). Consequently, only 8 environmental variables (alt, bio1,

Variables	Variable type	Min	1st Qu	Median	Mean	3rd Qu	Max	SD
Latitude	Mean latitude	21.90	28.45	31.10	34.42	41.67	48.90	7.90
Longitude	Mean Longitude	81.20	103.00	114.40	111.60	118.40	132.60	10.47
alt	Elevation of the site	2.00	23.50	144.00	911.70	1773.00	4947.00	1154.56
area	Surface area of the lake	0.01	6.72	27.95	225.80	122.33	4930.00	544.82
bio1	Annual mean temperature	-5.00	50.00	140.00	113.70	170.00	217.00	61.45
bio2	Mean diurnal range: mean of monthly (max temp–min temp)	70.00	82.00	110.00	104.50	118.00	146.00	19.95
bio3	Isothermality: (Bio2/Bio7) *100	21.00	24.00	25.00	30.16	36.50	53.00	9.78
bio4	Temperature seasonality (SD *100)	3242.00	6903.00	8807.00	9305.00	12168.00	15770.00	3468.65
bio5	Maximum temperature of warmest month	130.00	258.00	286.00	284.40	323.50	340.00	46.71
bio6	Minimum temperature of coldest month	-295.00	-196.50	-22.00	-79.59	9.50	93.00	109.16
bio7	Temperature annual range (Bio5–Bio6)	212.00	306.50	331.00	364.00	452.50	561.00	103.52
bio8	Mean temperature of wettest quarter	78.00	199.50	216.00	214.30	244.50	283.00	40.57
bio9	Mean temperature of driest quarter	-208.00	-117.00	50.00	-2.47	79.50	162.00	104.59
bio10	Mean temperature of warmest quarter	78.00	202.00	221.00	227.00	272.00	286.00	48.12
bio11	Mean temperature of coldest quarter	-213.00	-119.50	43.00	-12.32	57.00	149.00	99.40
bio12	Annual precipitation	82.00	418.50	919.00	806.30	1177.50	1480.00	409.61

bio13	Precipitation of wettest month	18.00	132.50	186.00	168.60	225.00	305.00	63.03
bio14	Precipitation of driest month	0.00	2.00	9.00	15.04	31.50	44.00	14.62
bio15	Precipitation seasonality(coefficient of variation)	39.00	55.00	86.00	83.35	106.50	138.00	26.64
bio16	Precipitation of wettest quarter	47.00	291.00	461.00	423.70	573.00	815.00	171.80
bio17	Precipitation of driest quarter	2.00	8.00	38.00	56.64	122.00	160.00	55.12
bio18	Precipitation of warmest quarter	47.00	291.00	443.00	403.20	516.00	676.00	157.02
bio19	Precipitation of coldest quarter	2.00	8.00	39.00	60.01	129.50	175.00	59.59

Tab.1 Comprehensive description of all the variables

***Min:** the smallest observation**;**

1st Qu: a value that separates the largest 75% of the observations from the smallest 25%;

Median: the median value of all the observations;

Mean: the mean value of all the observations;

3rd Qu: a value that separates the largest 25% of the observations from the smallest 75%;

Maximum: the largest observation;

SD: The standard deviation.

bio2, bio3, bio7, bio5, bio14, bio19; see details in Table 1) were contained in the analysis.

Ensemble modelling of species distributions

A total of nine different statistical models, contained: GLM, GAM, MARS, CTA, RF, MDA, ANN, GBM, SRE (Surface Range Envelop or usually called BIOCLIM, Busby 1991), were used to predict the presence-absence data of each fish species in the assemblage. All the models were performed in R (R Development Core Team 2013) with the package of “biomod2” (Thuiller et al., 2003). For each of the 92 species, nine models were constructed respectively with a random 70% subset, and the remaining 30% dataset were used to evaluate the model performance. The split-sample procedure was repeated 100 times for each species.

Model evaluation

The predictive model performance were evaluated using three main kinds of accuracy measures: Area under the receiver operating characteristic curve (AUC; Fielding and Bell, 1997), Cohen’s Kappa (Cohen, 1960) and True skill statistic (TSS ; Bohning et al., 2008).

Area under the receiver operating characteristic curve (AUC) is one kind of threshold-independent measures, A ROC (receiver operating characteristic curve) plot is obtained by plotting all sensitivity values (true positive fraction) on the y axis against their equivalent (1-specificity) values (false positive fraction) for all available thresholds on the x axis. The area under the ROC function (AUC) is usually taken to be an important index because it provides a single measure of overall accuracy that is not dependent upon a particular threshold (Deleo 1993; Fielding and Bell, 1997). The value of the AUC ranged from 0 to 1, with a score of 1 indicating a perfect fit, scores from 0.9 to 1 representing an excellent fit, scores from 0.8 to 0.9 representing good fit, and those scores from 0.5 to 0.6 implying discrimination that was no better than random (Swets, 1988).

Cohen's kappa (1960) has been adopted to alleviate the problem of overestimating accuracy. It measures the extent to which the agreement between observed and predicted is higher than that expected by chance alone. This statistic is used to assess inter-rater reliability when observing or otherwise coding qualitative / categorical variables. Kappa is considered to be an improvement over using % agreement to evaluate this type of reliability. Kappa has a range from 0-1.00, with larger values indicating better reliability.

TSS (True skill statistic) is defined as the average of the net prediction success rate for presence sites and that for absence sites. It has gained considerable theoretical interest over many years (Bohning et al., 2008), and it is considered the best available summary measure of model performance in medical diagnostic tests by some researchers (Biggerstaff, 2000). TSS takes into account both omission and commission errors, and success as a result of random guessing, and the values range from −1 to 1, where 1 indicates perfect agreement and values of zero or less indicate a performance no better than random. This index is closely related to the arithmetic mean of sensitivity and specificity. Overall accuracy (OA), defined as the probability that a site (either presence or absence) is correctly predicted, is the most common measure used in various disciplines including ecology.

MANOVA (Multivariate Analysis Of Variance) were conducted to explore the potential impacts of different evaluation methods on model performance.

Species characteristics definition

For each species predicted, species prevalence was defined as the fish occurrence rate in all the studied lakes. In addition, three main environmental ranges were determined and calculated for altitude, temperature and precipitation. The species altitude range was expressed as the difference between the average of the 10% highest altitude values and the 10% lowest altitude values among all the sites which species occurred. Temperature range described as the difference between the average of the 10% highest annual temperature and the 10% lowest annual temperature. While the precipitation range was defined as the difference between the average of the 10% highest annual mean precipitation and the 10% lowest mean precipitation among the sites which species occurred, respectively.

To investigate the observed difference between model performances and species characteristics, generalized liner model (GLM) was therefore applied.

Results

Overall, the nine SDMs performed well and showed good capacity on species prediction, as the three different evaluation methods (AUC, Kappa, TSS) and overall accuracy showed significantly high values, with the median values of AUC ranging from 0.680 to 0.891, overall accuracy ranging from 0.85 to 0.90, Kappa ranging from 0.358 to 0.658, TSS values ranging from 0.360 to 0.752 (Fig. 2). Among the entire model techniques, RF always yield the best models since the median AUC value is 0.891, median accuracy score is 0.90, median Kappa value is 0.658 and median TSS value is 0.752, all the values are the highest. Followed by GBM, FDA and GLM respectively, while SRE always gives the relative worst predictive performance with median AUC value of 0.680, accuracy value of 0.85, Kappa value of 0.358 and TSS value of 0.360, which showed the relative lowest values among all the models (Fig. 2). It is worth noting that even if the RF model which performs the best couldn’t give the equal and best predictions for all species, while SRE could also yield very good quality models for some species.

Results from the MANOVA showed that there is no significant difference between the three evaluations used in measuring the performance of nine models ($p>0.01$).

In order to explore the relationship between species prevalence and model performance, outcomes of all the nine models were taken into account. In the context, only the results of RF model were presented in Fig. 4. Among all the three evaluations, there is a nonlinear relationship between model performance and species prevalence.

The species prevalence around 30% could yield more accurately predictive models, while species prevalence below 30%, model performance showed slightly positive relationship with species prevalence (Fig.4).

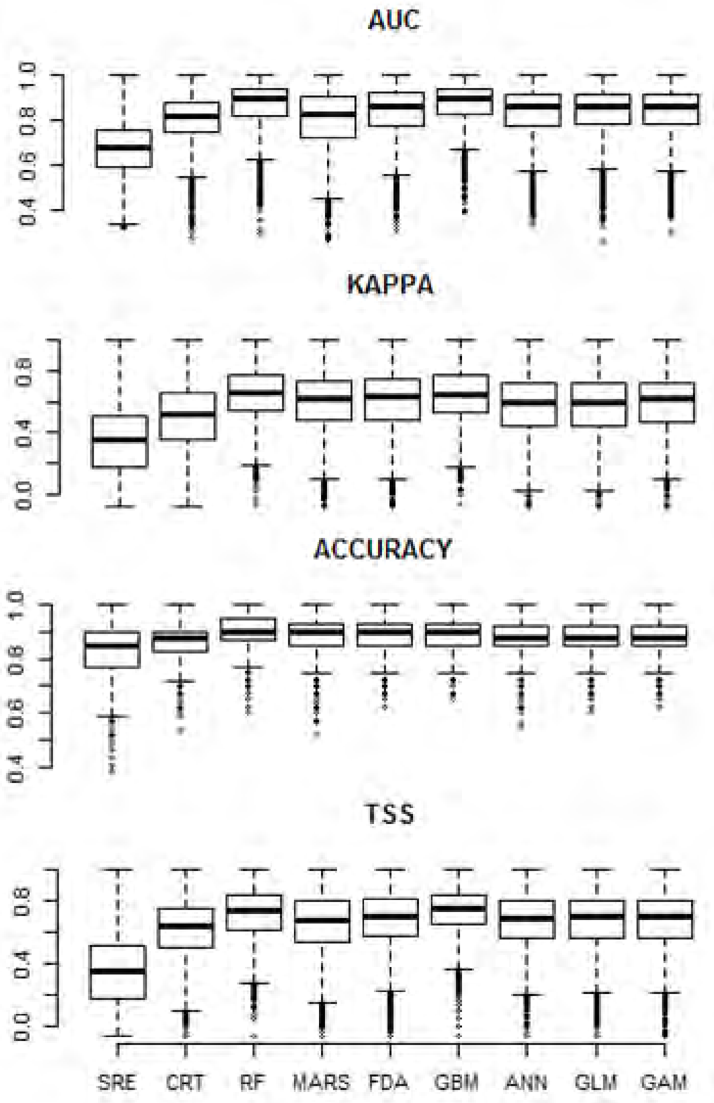


Fig.2 Performance of different SDMs

AUC: Area under the receiver operating characteristic curve; KAPPA: Cohen’s Kappa; TSS; True skill statistic; ACCURACY: Overall accuracy.

SRE (Surface Range Envelop); CTA (Classification Tree Analysis) ;MARS(Multivariate Adaptive Regression Splines) ; RF (Random Forest) ;FDA (Mixture Discriminant Analysis); GBM (Generalized Boosted Models) ; ANN (Artificial Neural Networks); GLM (Generalized Linear Models); GAM (Generalized Additive Models)

Here in our study, we determined three kinds of ecological range size, altitude, temperature and precipitation ranges. Relationships between species prevalence and species range size were found to be all positively related with all the three environmental range size (Fig.3).

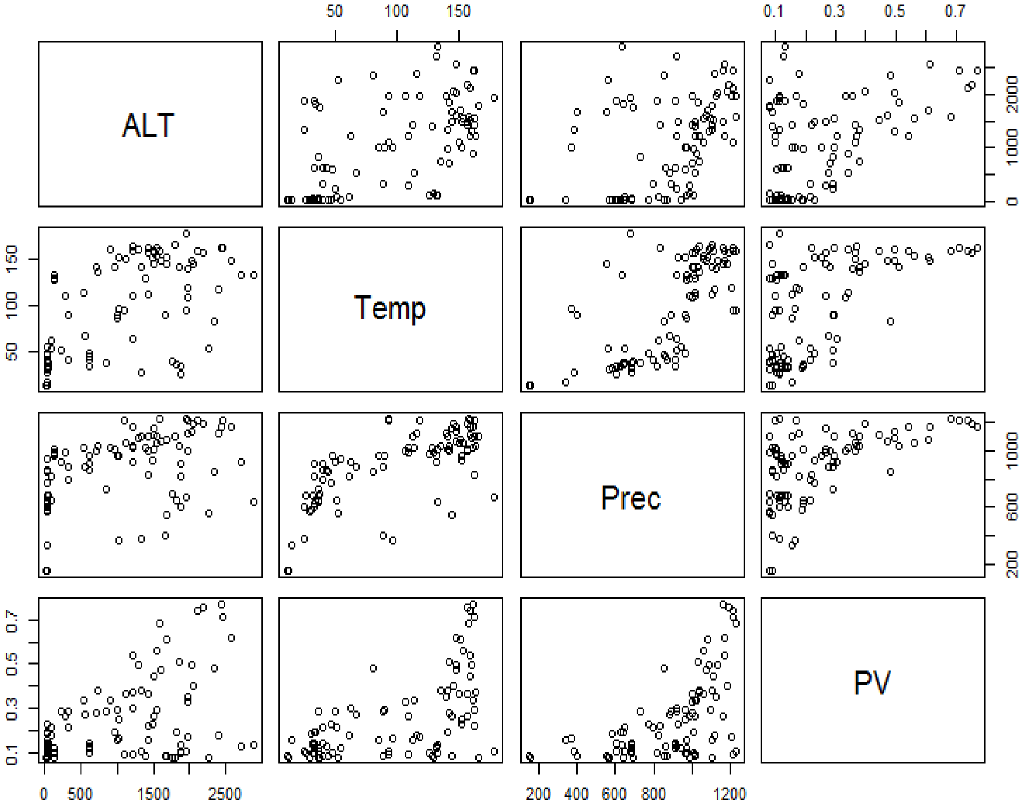


Fig.3 Relationships between species prevalence and three environmental range size

However, when take into account the species ecological range size, the altitude range size and precipitation range size were significantly negative related with the model performance, while unfortunately, temperature range size were found not significantly influence the model performance. Besides, performance of FDA and SRE seems only affected by species prevalence (Tab.2), but not correlated with species characteristics like temperature range size, precipitation range size and altitude range size. Moreover, influence from the uncertainty source on each of the nine models was also showed difference.

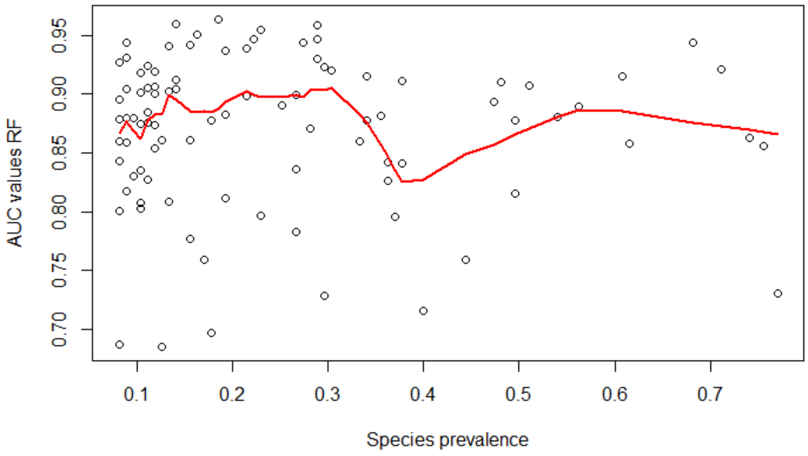
Tab. 2 Relations between model techniques and species characteristics

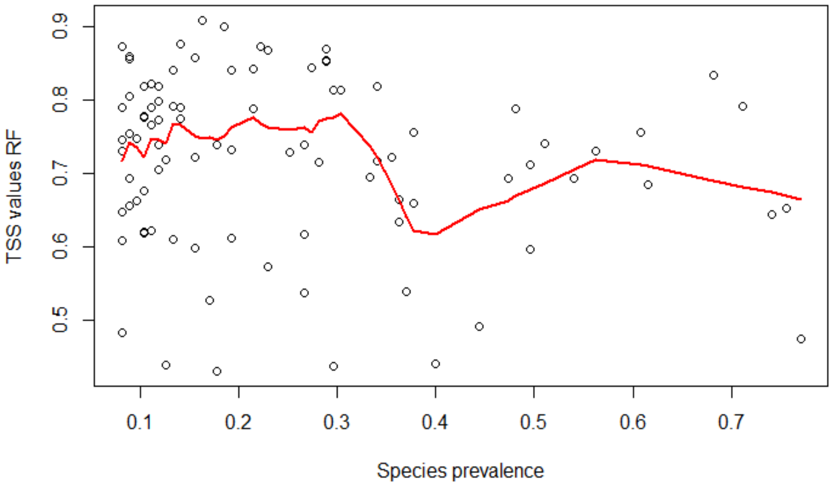
	ALTR		TR	PR		PV	
SRE AUC	NS		NS	NS		0.208	***
CTA AUC	-3.057E-05	***	NS	-1.03E-04	**	0.193	***
MARS AUC	-2.42E-05	**	NS	-9.72E-05	**	0.185	***
RF AUC	-3.43E-05	***	NS	-1.16E-05	***	0.172	***
FDA AUC	NS		NS	-1.28E-04	**	0.233	***
GBM AUC	-3.08E-05	***	NS	1.32E-04	***	0.178	***
ANN AUC	-4.41E-05	***	NS	-1.12E-04	***	0.206	***
GLM AUC	-3.43E-05	***	NS	-6.04E-05	*	0.253	***
GAM AUC	-3.87E-05	***	NS	-7.89E-05	*	0.192	***
SRE KAP	NS		NS	-1.93E-04	*	0.419	***
CTA KAP	-4.43E-05	*	NS	-2.12E-04	*	0.614	***
MARS KAP	-3.73E-05	*	NS	-1.52E-04	*	0.344	**
RF KAP	-3.62E-05	*	NS	-1.73E-04	*	0.283	***
FDA KAP	NS		NS	-1.81E-04	*	0.371	***
GBM KAP	-4.00E-05	*	NS	-2.26E-04	**	0.435	***
ANN KAP	-6.09E-05	***	NS	-1.74E-04	*	0.49	***
GLM KAP	-4.94E-05	**	NS	-1.44E-04	*	0.63	***
GAM KAP	-5.32E-05	**	NS	-1.95E-04	*	0.4399	***
SRE TSS	NS		NS	NS		0.3666	***
CTA TSS	-5.55E-05	***	NS	-2.01E-04	**	0.3626	***
MARS TSS	-4.25E-05	**	NS	-1.78E-04	**	0.268	***
RF TSS	-5.73E-05	***	NS	-2.23E-04	***	0.229	***
FDA TSS	NS		NS	-2.19E-04	***	0.261	***
GBM TSS	-5.20E-05	***	NS	-2.16E-04	***	0.209	***
ANN TSS	-7.89E-05	***	NS	-1.81E-04	***	0.307	***
GLM TSS	-5.66E-05	***	NS	-1.04E-04	*	0.382	***
GAM TSS	-5.81E-05	***	NS	-1.25E-04	*	0.268	***

The model techniques: SRE (Surface Range Envelop); CTA (Classification Tree Analysis) ;MARS(Multivariate Adaptive Regression Splines) ; RF (Random Forest) ;FDA (Mixture Discriminant Analysis); GBM (Generalized Boosted Models) ; ANN (Artificial Neural Networks); GLM (Generalized Linear Models); GAM (Generalized Additive Models) ;

The evaluation techniques: AUC (Area Under the receiver operating characteristic Curve); KAP(Cohen's Kappa); TSS (True skill statistic);

The species characteristics: ALTR (Altitude); TR(Temperature Range); PR(Precipitation Range); PV(Prevalence); **NS:** Not Significant





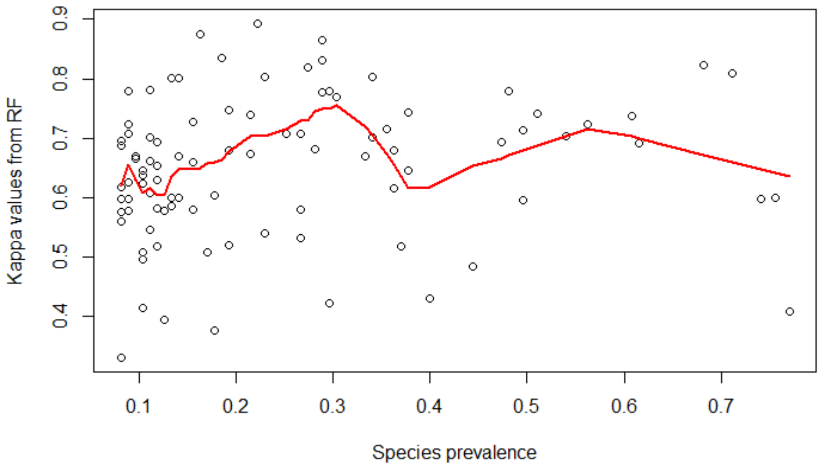


Fig.4 Relationships between species prevalence and model performance

Discussion

It is important to understand why and how species distribution models perform differently for different species before using the model predictions to make conservation decisions. Up until now, our research is the first study which quantified the uncertainties and impacts comprehensively from model techniques, environmental range size and species prevalence on the SDM performance in lake ecosystems in so large geographical scale. Overall, our results have compared the performance of nine widely used species distribution models; clarified that evaluation measures won't influence the model outcome; confirmed that species prevalence and environmental range size can strongly affect model performance and ecological predictions; and verified the hypothesis that specialized species could be better predicted than generalized species is plausible in aquatic ecosystems.

Among all the statistical techniques, RF was found to be the most reliable model for species prediction, while SRE predicted the worst. However, the predictions from different models varied a lot, even if for one given species, outcomes of prediction may vary from model to model. In view of that each predicts models relied on different mathematical functions, SDM will give the variety of results without doubt. Up until now, numerous of studies have compared the accuracy and performance of the predictions from different statistical techniques (Elith et al., 2007), and confirmed that results derived from different model techniques or different model-building assumptions can occasionally differ grossly (Thuiller, 2003; Luoto et al., 2004). Nevertheless, it was not so surprise since RF model gives the predictions by generating thousands of trees and aggregated with an average (Breiman, 2001), and the algorithm allow the model to avoid over-fit, this procedure could improve the predictive performance and reduce the variance (Elith et al., 2008). Thus, RF could be a robust technical modeling for species distribution prediction (He et al., 2010; Cheng et al., 2012; Grenouillet et al., 2011). Actually, plenty of publications have noted the algorithm which Random Forest relied on, they thus present the ensemble modelling framework which aggregated several single models and given the average or consensus results (Araujo and New, 2007). Several former studies have verified that among plenty of mathematical models, only RF could show the equal performance with the average outcomes of several model techniques. Therefore ensemble modelling was also regarded as the best solutions to reduce the single model uncertainties and bias (Grenouillet et al., 2011; Buisson et al., 2010b).

Considered that some authors argued that AUC statistic may be biased for the species that occupy a small proportion of the study area (Lobo et al., 2008), in current research, three different measurements have been adopted in order to reduce the potential bias may emerge from single measurement. However, in current study, among all the evaluation results from AUC, TSS and Kappa, there is no significant difference between the three evaluate methods, which interpret that the use of evaluation measures didn’t interrupt the outcomes of statistical models, we can thus verified that the relationship between species characteristics and model performance was not artificial associated with use of accuracy measures (Newbold et al., 2009).

In some previous studies, publicized effects of species prevalence on model performance showed complicated with both positive and negative relationships (Luoto et al., 2005; Brotons et al., 2004; Mantel et al., 2001). In current study, we found that species with prevalence around 30% could yield more accurate model performance, however, the relationships between species prevalence and model performance was nonlinear, model performance increased slightly with species prevalence values below 0.3. Actually, these results were completely in consistent with the former research which argued that species with high prevalence will occupy large areas with variety of habitat environments which would produce more overall errors (Segurado and Araujo, 2004; Luoto et al., 2005). Actually, species prevalence has long been reported which will affect model performance, however few ecologist take into account the species prevalence when they make the species distribution modeling (Fielding and Bell, 1997). Manel et al. (2001) revealed that model performance were highly associated with species prevalence and thus recommend that species distribution model with presence-absence data should take more attention of species prevalence.

Overall in our study, the performance and ecological predictions of the species distribution modelling was negatively associated with geographical and environmental range size, such as altitude range size and precipitation range size. Actually, A general pattern has been widely acknowledged is that species with smaller geographical extent and strict ecological requirements (i.e. specialists) yield models with higher accuracy than those with larger areas of occupancy (i.e. generalists), this hypothesis have been verified in various of ecosystems, among the species as butterfly, insect, reptile, bird, and mammal (Stockwell and Peterson, 2002; Brotons et al, 2004; Segurado and Araujo 2004; Kadmon et al., 2003; Seggurado and Araujo, 2004; Hernandez et al., 2006; Tsoar et al., 2007; Franklin et al., 2009). Our research has verified this hypothesis to be plausible to fish species in lake ecosystems. To our knowledge, the current research is the first research which takes into account fish species in lakes in so large scale, therefore, it would not only contribute to understand the uncertainty of species distribution models, but also help to improve the quality of fish species prediction which will ultimate benefit to the fish biodiversity conservation and biodiversity management all over the world.

When considering about the reason why specialized species could yield more accurate model prediction than generalized one, some authors argued that species with larger distribution area contained discrete populations that show different response to the environment and thus decrease the accuracy of the model prediction (Stockwell and Peterson, 2002; Brotons et al., 2004). Some others explained these as the difference of species' niche width, since former studies have found that species with narrow, well-defined niches which also have better-defined climate and habitat requirements could be better modeled than those with broader niches (Boone and Krohn, 1999; Pearce et al., 2001; Kadmon et al., 2003). But our data couldn't well support and explain the species niche hypothesis, further study should take more factors into account in order to better understand the mechanism of uncertainty in species distribution models from species attributes and characteristics. Last but not least, McPherson and Jetz (2007) attributed the effects of species' ecological characteristic on SDMs were influenced by the available data quality or by making it difficult to statistically capture the relationship between the species distribution and environmental conditions. Following these hypothesis, we can conclude that endemic species were modeled more accurately than non-endemic species. Given all of these, greater consideration should therefore be taken on the predictions of fish species with larger elevation range size and precipitation range size in China under the future impacts of global climate change, since the projections may somehow unreliable for the conservation and management purpose. On contrary, ecological predictions for the endemic fish species which showed better defined climate and habitat requirements should be more accurately. These results will benefit a lot to the conservation of biodiversity for fish species in China, since there are numerous of endemic and specialized fish species lived in the lakes across China, and urgent stages were planned to conserve the fish species based on the prediction results (He et al., 2010). However, it is worth noting that in our study, temperature range size didn't show a significantly affect on the model performance, this founding contradicted some former results in fish species (Grenouillet et al., 2011). In the author's opinion, we contributed this to the large range size of the geographical factor, and the effects of the temperature range may be hidden since that altitude is a complicated factor.

To conclude, we strongly recommend paying more attention on the following aspects when predicting the potential impacts of global climate change on fish biodiversity and distribution using species distribution models. First of all, more suitable model should be implemented and more statistical model techniques should be taken into account in the ensemble model framework; Second, high quality fish presence-absence dataset should be compiled; Last but not least, species ecological and geographical characteristic should be taken into account when predicting the species distribution and diversity. Following these recommendations, predictions and projections of fish species distribution based on the results of SDM could be more reliable and therefore robust enough for management and conservation decisions made under the future impacts of global climate change.

Acknowledgements

This research was financially supported by Grant Nos. 200903048 and 20130305 from the “Special Fund for Agro-scientific Research in the Public Interest” of China, Nos. 30830025 and 30900182 from the National Natural Science Foundation of China. We are grateful to the China Scholarship Council (CSC) for their financial support of Chuanbo Guo’s study abroad.

Reference

Araujo, M. B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. J. Biogeogr. 33(10), 1677-1688.

Araújo, M. B., New, M., 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. 22(1), 42-47.

Araújo, M. B., Rahbek, C.,2006. How does climate change affect biodiversity?

Sci.N.Y. Then W., 313(5792), 1396.

Barbet-Massin, M., Walther, B. A., Thuiller, W., Rahbek, C., Jiguet, F., 2009. Potential impacts of climate change on the winter distribution of Afro-Palaearctic migrant passerines. Biol. Letters 5(2), 248-251.

Biggerstaff, B. J., 2000. Comparing diagnostic tests: a simple graphic using likelihood ratios. Stat. Med. 19(5), 649-663.

Boone, R. B., Krohn, W. B., 1999. Modeling the occurrence of bird species: are the errors predictable? Ecol. Appl. 9(3), 835-848.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24(2), 123-140.

Breiman, L., 2001. Random forests. Mach. learn. 45(1), 5-32.

Brotons, L., Thuiller, W., Araújo, M. B., Hirzel, A. H., 2004. Presence‐absence versus presence‐only modelling methods for predicting bird habitat suitability. Ecography 27(4), 437-448.

Brummer, T. J., Maxwell, B. D., Higgs, M. D., Rew, L. J., 2013. Implementing and interpreting local‐scale invasive species distribution models. Divers.Distrib. doi: 10.1111/ddi.12043.

Buisson, L., Grenouillet, G., 2009. Contrasted impacts of climate change on stream fish assemblages along an environmental gradient. Divers. Distrib. 15(4), 613-626.

Buisson, L., Thuiller, W., Casajus, N., Lek, S., Grenouillet, G., 2010. Uncertainty in ensemble forecasting of species distribution. Global Change Biol. 16(4), 1145-1157.

Buisson, L., Thuiller, W., Lek, S., Lim, P., Grenouillet, G., 2008. Climate change hastens the turnover of stream fish assemblages. Global Change Biol.14(10), 2232-2248.

Cawsey, E. M., Austin, M. P., & Baker, B. L.,2002. Regional vegetation mapping in Australia: a case study in the practical use of statistical modelling. Biodivers. Conserv. 11(12), 2239-2274.

Cheng, L., Lek, S., Lek-Ang, S., Li, Z., 2012. Predicting fish assemblages and diversity in shallow lakes in the Yangtze River basin. Limnologica 42(2), 127-136.

Cohen, J.,1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20(1), 37-46.

De'Ath, G., 2002. Multivariate regression trees: a new technique for modeling species-environment relationships. Ecology 83(4), 1105-1117.

Elith, J., Leathwick, J., 2007. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. Divers. Distrib. 13(3), 265-275.

Elith, J., Leathwick, J. R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77(4), 802-813.

Fielding, A. H., Bell, J. F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24(1), 38-49.

Friedman, M., 1991. The Haredi (Ultra-Orthodox) Society—Sources, Trends and Processes. Jerusalem: The Jerusalem Institute for Israel Studies.

Grenouillet, G., Buisson, L., Casajus, N. Lek, S.,2011. Ensemble modelling of species distribution: the effects of geographical and environmental ranges. Ecography 34(1): 9-17.

Guisan, A., Zimmermann, N. E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135(2), 147-186.

Guisan, A., Graham, C. H., Elith, J., Huettmann, F., 2007. Sensitivity of predictive species distribution models to change in grain size. Divers. Distrib. 13(3), 332-340.

Hastie, T. J., Tibshirani, R. J., 1990. Generalized Adolitive Models (No. 43). CRC Press.

He, Y., Wang, J., Lek-Ang, S., Lek, S., 2010. Predicting assemblages and species richness of endemic fish in the upper Yangtze River. Sci. Total Environ. 408(19), 4211-4220.

Hernandez, P. A., Graham, C. H., Master, L. L., & Albert, D. L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29(5), 773-785.

Hijmans, R. J., Cameron S. E., Parra J. L., Jones P. G., Jarvis A., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol.25, 1965–1978.

Hutchinson, M. F., 1995. Interpolating mean rainfall using thin plate smoothing splines. Int. J. Geogr. Inf. Sci. 9(4), 385-403.

Kadmon, R., Farber, O., Danin, A.,2003. A systematic analysis of factors affecting the performance of climatic envelope models. Ecol. Appl. 13(3), 853-867.

Lobo, J. M., Jiménez ‐ Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecol. Biogeogr. 17(2), 145-151.

Luoto, M., Pöyry, J., Heikkinen, R. K., Saarinen, K., 2005. Uncertainty of bioclimate envelope models based on the geographical distribution of species. Global Ecol. Biogeogr. 14(6), 575-584.

Manel, S., Williams, H. C., Ormerod, S. J., 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. J. Appl. Ecol. 38(5), 921-931.

Marmion, M., Luoto, M., Heikkinen, R. K., Thuiller, W., 2009b. The performance of state-of-the-art modelling techniques depends on geographical distribution of species. Ecol. Model. 220(24), 3512-3520.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., Thuiller, W., 2009a. Evaluation of consensus methods in predictive species distribution modelling. Divers. Distrib. 15(1), 59-69.

McCullagh, P., Nelder, J. A.,1989. Generalized linear models (Monographs on statistics and applied probability 37). Chapman Hall, London.

McPherson, J., Jetz, W., 2007. Effects of species' ecology on the accuracy of distribution models. Ecography 30(1), 135-151.

Park Y.S., Grenouillet G., Esperance B., Lek S., 2006. Stream fish assemblages and basin land cover in a river network. Sci. Total Environ. 365: 140-153.

Pearce, J., Ferrier S., 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecol. Model. 133: 225-245.

Phillips, S. J., Anderson, R. P., Schapire, R. E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190(3), 231-259.

R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Ripley, B. D.,1996. Pattern recognition and neural networks. Cambridge Uni. Press, Cambridge.

Roura-Pascual N., Brotons, L., Peterson, A. T., Thuiller, W., 2009. Consensual predictions of potential distributional areas for invasive species: a case study of Argentine ants in the Iberian Peninsula. Biol. Invasions. 11(4), 1017-1031.

Segurado, P., Araujo, M. B., 2004. An evaluation of methods for modelling species distributions. J. Biogeogr. 31(10), 1555-1568.

Stockwell, D. R., Peterson, A. T.,2002. Effects of sample size on accuracy of species distribution models. Ecol. Model. 148(1), 1-13.

Thuiller, W., 2003. BIOMOD–optimizing predictions of species distributions and projecting potential future shifts under global change. Global Change Biol. 9(10), 1353-1362.

Thuiller, W., Lavorel, S., Araújo, M. B., Sykes, M. T., Prentice, I. C., 2005. Climate change threats to plant diversity in Europe. P. Natl. Acad. Sci. USA. 102(23), 8245-8250.

Tisseuil, C., Leprieur, F., Grenouillet, G., Vrac, M., Lek, S., 2012. Projected impacts of climate change on spatio‐temporal patterns of freshwater fish beta diversity: a deconstructing approach. Global Ecol. Biogeogr. 21(12), 1213-1222.

Tsoar, A., Allouche, O., Steinitz, O., Rotem, D., Kadmon, R., 2007. A comparative evaluation of presence‐only methods for modelling species distribution. Divers. Distrib. 13(4), 397-405.

Wang, S.M., Dou, H.S., 1998. Lakes of China. Science Press, Beijing.

Whittaker, R. J., Nogués‐Bravo, D., Araújo, M. B., 2007. Geographical gradients of species richness: a test of the water - energy conjecture of Hawkins et al.(2003) using European data for five taxa. Global Ecol Biogeogr. 16(1), 76-89.

Wu, X.W., 1964. The Cyprinoid Fishes of China, Vol. 1. People’s Press, Beijing (in Chinese).

Wu, X.W., 1977. The Cyprinoid Fishes of China, Vol. 2. People’s Press, Beijing (in Chinese).

Yu, D., Chen, M., Zhou, Z., Eric, R., Tang, Q., Liu, H., 2013. Global climate change will severely decrease potential distribution of the East Asian coldwater fish Rhynchocypris oxycephalus (Actinopterygii, Cyprinidae). Hydrobiologia. 700(1), 23-32.

Zhao, S., Fang, J., Peng, C., Tang, Z., Piao, S., 2006. Patterns of fish species richness in China's lakes. Global Ecology Biogeogr. 15(4), 386-394.