



OPEN Integrated analytical framework for identifying factors related to the ecological degradation of lakes

Yong Zeng¹✉, Yanwei Zhao² & Wei Yang²

The causal relationships between external driving forces and the ecological degradation of lakes are characterized as complex and multidimensional, with multiple inputs and outputs, nonlinearity, and many interactions. Conventional parametric statistical methods such as correlation analysis and multiple linear regression cannot handle these characteristics simultaneously. Thus, we developed an integrated analytical framework to screen, identify, and predict the factors related to the ecological degradation of lakes based on redundancy analysis (RDA), variance partitioning analysis (VPA), and principal component analysis-based generalized additive models (PCA-based GAM). The RDA and VPA methods were employed to identify and rank the driving factors that explained the decrease in species richness (specifically of key aquatic organisms, including phytoplankton, submerged plants, zooplankton, benthic animals, and fish), which is a critical ecological indicator closely associated with lake ecological degradation. PCA-based GAM was used to explore the patterns associated with driving forces. The driving forces related to the changes in species richness during the 35 years from 1986 to 2020 were investigated in Baiyangdian (BYD) Lake, China. Three categories of driving forces were identified: anthropogenic pollution, climate change, and hydrological conditions. Significant detrimental changes in species richness were detected in the first decade, followed by relative stability in the next decade, and favorable changes since 2015. Anthropogenic pollution, climate change, and hydrological conditions explained 41%, 18%, and 13% of the total variance, respectively. The best predictive model structures included the water level (WL), air temperature (AT), total phosphorus (TP), and (WL*TP) interaction, and they explained 98.4% of the total data variance. The proposed method offers actionable solutions for lake management, including real-time ecological health monitoring, adaptive strategies and indicating ecological degradation.

Keywords Baiyangdian lake, Driving force, Generalized additive model, Redundancy analysis, Species richness

In recent years, the ecological health and sustainability of lakes have been severely threatened by rapid urbanization, global climate change, and intense anthropogenic activities¹. These pressures have led to the widespread degradation of lake ecosystems worldwide, including harmful algal blooms, enhanced hypolimnion anoxia, aquatic biodiversity losses, and abrupt ecological shifts². Identifying the key factors that influence the ecological structure and functions of lakes is crucial for understanding the ecological responses to multiple stressors, thereby facilitating effective ecological restoration strategies.

Changes in the ecological structure and functions of lakes are complex multivariate processes driven by multiple factors, including natural and human-induced factors^{3,4}. However, previous studies mainly focused on the effects of driving forces on some fragmented elements of lake ecosystems, such as the driving forces that affect water quality^{5–8}, nutrient loads⁹, the richness of certain species^{10–13}, and dynamics of certain communities^{14–17}. Some studies focused on how hydrological regimes affect the response in terms of the lake health index^{18–20}, such as species biomass and structure, eco-exergy, and structural eco-exergy. Much progress has been made in identifying factors that affect the fragmented characteristics of ecosystems. However, further research is required to understand how multiple stressors drive ecological changes due to the lack of long-term records regarding hydrology, water chemistry, and biota data, and novel multivariate statistical analysis methods^{21,22}.

¹State Key Laboratory of Heavy Oil Processing, Beijing Key Laboratory of Oil & Gas Pollution Control, College of Chemical Engineering and Environment, China University of Petroleum, Beijing 102249, China. ²State Key Laboratory of Water Environment Simulation, School of Environment, Beijing Normal University, Beijing 100875, China. ✉email: yongzeng1974@163.com

Previous quantitative studies have not systematically explored how key factors influence holistic ecosystems in the long term. Therefore, this study aims to unravel the multi-dimensional driving mechanisms underlying the ecological degradation of Baiyangdian (BYD) Lake, China, focusing on the following scientific inquiries: What natural and anthropogenic factors (climate, hydrology, pollution) are the primary drivers of declining species richness (specifically of key aquatic organisms, including phytoplankton, submerged plants, zooplankton, benthic animals, and fish) in BYD Lake? How do these drivers influence ecosystem states through nonlinear relationships and interactions? How can multivariate statistical methods be integrated to quantify the relative contributions of driving factors and develop predictive models? To address these challenges, the integrated framework of Redundancy Analysis (RDA), Variance Partitioning Analysis (VPA), and Principal Component Analysis-Generalized Additive Model (PCA-GAM) applied to BYD's long-term ecological data for the first time, identifies key drivers (RDA), quantifies their contributions (VPA), and establishes predictive models (GAM), addressing the limitations of single methods. Based on historical context and existing studies on BYD's ecological degradation, the following hypotheses are proposed: (1) Nutrient inputs (TP, TN) are the primary drivers of ecological degradation due to their direct role in eutrophication and suppression of aquatic biodiversity. (2) Climate change (rising temperatures) and hydrological fluctuations (declining water levels) indirectly exacerbate ecological stress by altering physicochemical conditions. (3) Species richness exhibits threshold effects in response to drivers.

In order to test the above research hypotheses, the following research were carried out, including: (1) identifying the driving forces related to the ecological degradation of BYD Lake; (2) quantifying the specific contributions and relative importance of these different driving forces; and (3) quantifying the patterns of ecological degradation in response to these driving forces.

Data and methods

Study area

BYD Lake (38.44°–38.59°N, 115.45°–116.06°E) located in Hebei Province, China, has a maximum area of 366 km² and mean depth of 2–3 m¹² (Fig. 1). BYD Lake is the largest shallow lake in the North China Plain region and it is important in terms of supplying water, commercial fishing, tourism, and recreation, as well as for the conservation of its wildlife and wetland vegetation²³. The lake is dominated by the warm temperate semi-arid continental monsoon climate, with an average annual temperature of 12.5 °C²⁰, average annual precipitation of about 497 mm, and annual evaporation of 1637 mm²⁴. Due to high-intensity human activities and climate change, inflow into the lake has decreased as a consequence of upstream dam interception and increased water consumption, thereby resulting in a sharp decrease in the water level of the lake and water quality^{25,26}. The water level decreased from 8.9 m in the 1950s to 6.9 m in the 2000s. At most monitoring sites, the water quality is class IV or V (a contamination level defined by environmental quality standards for surface water in China (GB 3838 – 2002))²⁷. Class IV water is slightly polluted and suitable for industrial water supply and non-contact recreational purposes. Class V water is moderately polluted, primarily intended for agricultural irrigation and general landscape water areas. BYD Lake is a typical macrophyte-dominated shallow lake and it is under pressure due to eutrophication, where the dominant plant composition has tended to shift from submerged species to

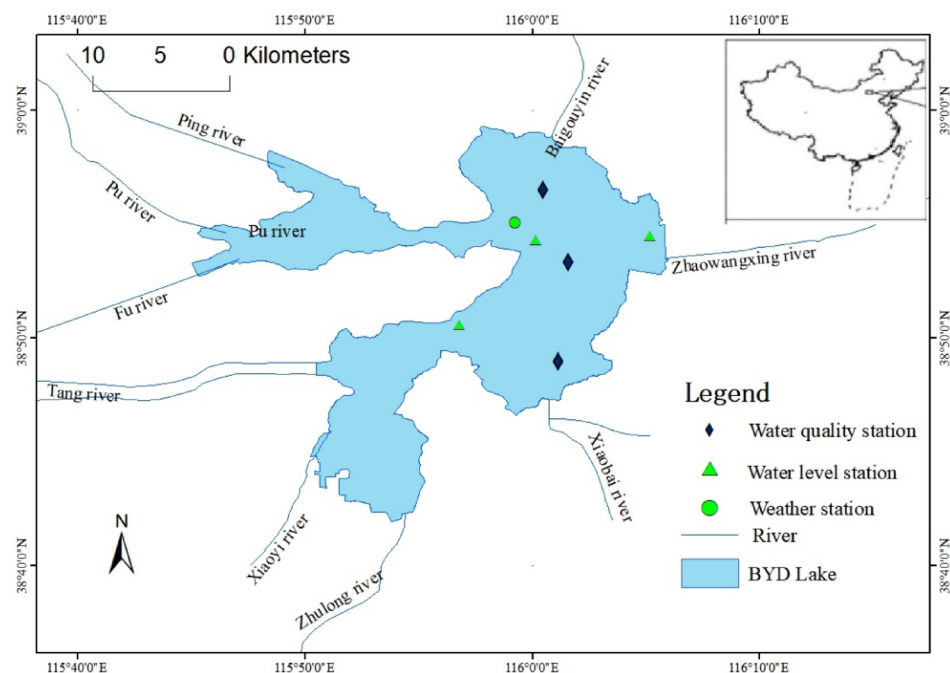


Fig. 1. Location of BYD Lake and distribution of sampling sites. Maps were created using ArcGIS 10.2 (Environmental Systems Research Institute, USA. <https://www.esri.com/>).

floating-leaved and emergent species¹¹. Since the 1990s, continuous decreases have occurred in the number of species of aquatic vegetation, phytoplankton, zooplankton, and macrobenthos^{28,29}.

Data sources

The data used for analysis were acquired from multiple sources. Meteorological data (from 1986 to 2020) were obtained from the HeHuaDaGuanYuan Meteorological Station (116°0′2.25″E, 38°55′12.86″N). Hydrological data (from 1986 to 2020) were provided by the Baoding Water Resources Bureau, Hebei Province (<http://slj.baoding.gov.cn/>), and water quality measurements (1986 to 2020) were obtained from data released by the Ministry of Ecology and Environment of China (<https://www.mee.gov.cn/>). The areas of submerged plants (1986–2020) were extracted from time-series remote sensing imagery in a study by Wang et al.³⁰. Diatom assemblage data (1986–2020), including the *Cyclotella meneghiniana* richness, were acquired from a previous sediment core analysis². Richness data for phytoplankton, zooplankton, benthic animals, and fish (1986–2020) were acquired from previous studies^{14,27,28,31}. Partial missing data were supplemented by using the linear interpolation method.

Methods

A flowchart illustrating the analytical methods used in this study is shown in Fig. 2. The entire process was divided into three parts. The part shown on the left in Fig. 2 involved collecting data from literature and screening indicators representing driving forces and ecosystem responses. The part shown on the middle in Fig. 2 involved the RDA-VPA-GAM analytical methods framework. The RDA, as a constrained ordination method, effectively analyzes relationships between multivariate environmental factors and biological communities, making it suitable for revealing complex ecosystem driving mechanisms³². The VPA addresses the limitation of traditional methods by quantifying the independent and interactive contributions of multi-source drivers³³. GAM's flexibility in handling nonlinearity³⁴ between ecosystem responses and driving forces. GAM captures these complexities through smoothing functions, outperforming traditional linear models. The GAM offers several advantages over multiple linear regression (MLR) and machine learning models. Unlike MLR, which assumes linear relationships, GAM captures nonlinearities through smooth functions, making it better suited for complex ecological data. GAM's interpretability is another strength, providing visual insights into the effect of each predictor on the response, unlike machine learning models that are often difficult to interpret. Additionally, GAM is less computationally demanding than many machine learning models, making it suitable for smaller datasets^{34,35}. Combined with PCA for dimensionality reduction, GAM integrates multi-dimensional biological responses into a composite index, resolving the high-dimensionality and collinearity of ecological data³⁵. The part shown on the right in Fig. 2 involved identifying the main driving forces using the RDA method, determining the relative contributions of the main driving forces with the VPA method, and establishing response patterns by using a principal component analysis (PCA)-based generalized additive model (GAM).

Screening indicators of driving forces

Changes in the ecosystem in BYD Lake appear to have occurred since the 1960s, and the water surface shrank dramatically after the construction of an upstream reservoir in the early 1960s³⁶. The multi-annual average

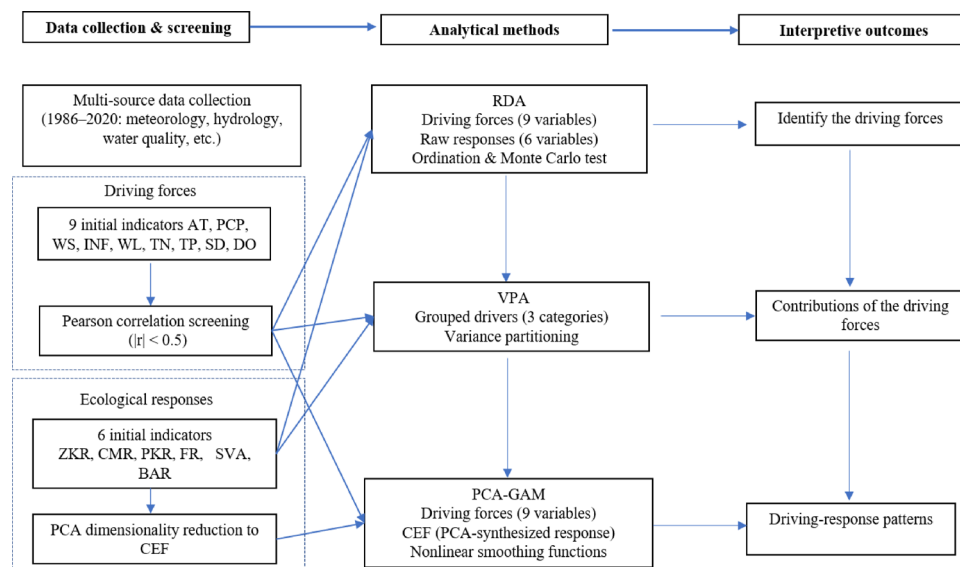


Fig. 2. A flowchart of the analytical frame used in this study (1) RDA-Redundancy Analysis; VPA-Variance Partitioning Analysis; PCA-Principal Components Analysis; GAM-Generalized Additive Model. (2) AT- air temperature; WL-water level; INF-inflow; WS- wind speed; PCP- precipitation; TN- total nitrogen; TP- total phosphorus; DO-dissolved oxygen; SD-secchi depth. (3) ZKR- zooplankton's richness; CMR- cyclotella meneghiniana's richness; PKR- phytoplankton's richness; FR- fish's richness; SVA- submerged vegetation areas; BAR- benthic animals' richness.

inflow into BYD Lake decreased from $1.94 \times 10^8 \text{ m}^3$ in the 1950s to $0.1 \times 10^8 \text{ m}^3$ in the 2000s¹⁹. Large amounts of nutrients have also been discharged into the lake, with multiyear averaged total nitrogen (TN) and total phosphorus (TP) loading amounts of 2018 t/a and 313 t/a, respectively³⁷. BYD Lake has become eutrophic due to the ecological threshold of nutrients has been exceeded since 2012³⁸. In addition, climate change has significantly influenced the water quality in BYD Lake, thereby resulting in the growth of phytoplankton and aggravated lake pollution³⁹. According to previous studies and the characteristics of the study area, nine indicators were selected to represent three types of driving forces: climatic change, anthropogenic pollution, and hydrological conditions, as shown in Table 1. The climatic change indicators were the air temperature (AT), precipitation (PCP), and wind speed. The indicators of anthropogenic pollution were total nitrogen, total phosphorus, Secchi depth, and dissolved oxygen. The indicators of hydrological conditions were lake inflow from rivers (INF) and WL. Detailed information about these indicators and the reasons for choosing them are shown in Table 1.

The collinearity among driving variables will weaken the explanatory power of the model. Therefore, Pearson correlation analysis is used for collinearity diagnosis, and the highly collinear driving variables are excluded. If $|r| > 0.5$ or 0.8 , it indicates a high degree of correlation between variables, and there may be a collinearity problem⁴⁰. We performed a collinearity analysis on the nine selected driving variables, as illustrated in the Fig. 3. The results show that all Pearson correlation coefficients are below 0.5 ⁴⁰, indicating that the selected indicators exhibit strong independence from one another.

Screening ecological response indicators

We utilized long-term data because some changes and related trends could be determined over longer time scales. Species richness is recognized as a critical issue regarding climate resilience because it enhances the stability of ecosystem functions and services in changing environments⁴¹. Species richness effectively indicates the characteristics of the community structure and functions, and thus it provides a measure of the ecosystem's state^{42,43}. In this study, six indicators were selected: zooplankton richness, *Cyclotella meneghiniana* richness, phytoplankton richness, fish richness, submerged vegetation areas, and benthic animal richness. These were integrated into a comprehensive index via principal component analysis (PCA) to characterize the holistic ecosystem response in Baiyangdian Lake.

RDA and VPA

RDA was applied to evaluate the effects of environmental variables on the ecosystem's state. RDA is a standard constrained ordination method that can be used to analyze the leading causes of variations in species richness by assessing the correlations between responses and explanatory variables^{44,45}. RDA can visually display species and rankings of environmental factors on a graph⁴⁵. The Monte Carlo permutation test was used to test the significance of the constraint ranking model⁴⁶.

CANOCO 5 software was used to conduct RDA. The nine indicators of driving forces were used as explanatory variables and the six ecosystem response indicators as response variables. The response data in this study were compositional and the gradient had a length of 0.8; therefore, a linear method is recommended³². The explanatory and response variables were both centered and standardized before RDA.

In addition, VPA was performed to explore and contrast the relative contributions of the explanatory variables³³. Three explanatory categories were assumed: climate change, hydrological conditions, and anthropogenic pollution (Table 1). The explanatory and response variables were $(\log(x + 1))$ transformed before VPA. The “Varpart ()” function in R software was employed to perform VPA.

GAM

GAM is a non-parametric extension of multiple linear models⁴⁷. The advantage of GAM is that it can directly fit the nonlinear relationships between the response variable and multiple explanatory variables³⁵. GAM is a strong model for identifying nonlinear relationships and offering interpretability, but whether it is the “best” model depends on the specific characteristics of the data and the research objectives. Cross-validation or performance metrics such as AIC or R^2 can help determine whether GAM is truly the best model for a specific

Driving forces	Indicators	Unit	Reasons
Climate change	Air temperature	°c	It is directly related to water temperature, the most crucial driving force on species' metabolism and dynamics.
	Precipitation	mm	It affects hydrology and water quality
	Wind speed	m/s	It affects the spatial distribution of plankton ¹⁴
Hydrological conditions ^a	Inflow	m^3/s	It affects the hydrological cycle, water quantity, water quality, and ecological balance ¹⁰ .
	Water level	m	It affects the distribution of submerged vegetation and benthos
Anthropogenic pollution ^b	Total nitrogen	mg/L	The main factor influencing species richness of phytoplankton, macrophytes, zooplankton, and benthos.
	Total phosphorus	mg/L	The main driving force of eutrophication is phosphorus, which is the limiting nutrient in BYD Lake ³⁷ .
	Secchi depth	cm	It affects the photosynthesis of aquatic plants
	Dissolved oxygen	mg/L	It affects the cycles of carbon, nitrogen, and phosphorus and maintains biological respiration and metabolism ¹⁶ .

Table 1. Screening for key indicators related to degradation of ecosystem in BYD Lake. **a.** Key factors related to hydrological conditions include dwindling water resources due to reservoir interception and excessive upstream water consumption **b.** Anthropogenic pollution can be classified as nutrient pollution, organic pollution, and light extinction

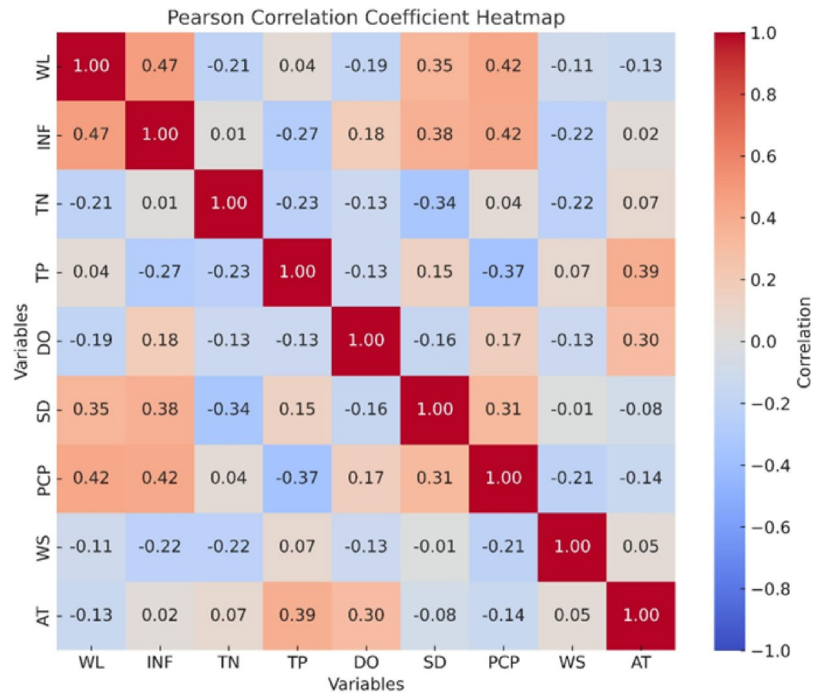


Fig. 3. Pearson correlation coefficient heatmap of driving indicators.

dataset or problem³⁴. GAM has been used widely in environmental science research to facilitate assessments of the nonlinear relationships among covariates^{48,49}. The general formula for GAM is⁴⁷:

$$g(E(Y)) = \beta_0 + f_1(X_1) + \dots + f_m(X_m) \quad (1)$$

where $E(Y)$ is the expectation of the response variable Y , $g(\cdot)$ is the link function, β_0 is the intercept, and $f_j(\cdot)$ is the smoothing function for the predictor variables X_j .

The GAM model is only applicable to a single response variable, so the six selected response indicators reflecting ecosystem attributes were first synthesized into a comprehensive indicator. Due to the correlations between ecological attributes, the six indicators had strong correlations and they could be reduced to one dimension by PCA. Assuming that there are n samples with p indexes, which are regarded as p random variables and recorded as X_1, X_2, \dots, X_p , by using PCA, p indicators are standardized, before performing linear regression to obtain k principal components $F_1, F_2, \dots, F_k (k \leq p)$ according to Eq. (2)⁵⁰:

$$\begin{aligned} F_1 &= \alpha_{11}ZX_1 + \alpha_{21}ZX_2 + \dots + \alpha_{p1}ZX_p \\ F_k &= \alpha_{1k}ZX_1 + \alpha_{2k}ZX_2 + \dots + \alpha_{pk}ZX_p \end{aligned} \quad (2)$$

where $\alpha = \frac{a}{\sqrt{\lambda}}$, a is the factor loading, λ is the eigenvalue, and ZX is the standardized index. The variance contribution of each principal component β_i was then used as a weight factor to construct a comprehensive evaluation function (CEF), as follows.

$$CEF = \beta_1F_1 + \beta_2F_2 + \dots + \beta_mF_m \quad (3)$$

PCA was performed with SPSS software. The GAM method was implemented using the “mgcv” package in R software³⁴. The CEF was entered as the response variable and the indicators of driving forces were treated as explanatory variables. The specific steps are described as follows. The first step involved detecting the significance of a single factor. The driving variables were $\log(x+1)$ transformed prior to GAM analysis, and then the nonlinear effects of each driving variable on the response variable were examined one by one. An indicator was retained if it was significant. The significance was then analyzed for the other indicators and they were gradually added to the GAM model. The next step involved analyzing the interaction terms in the model and determining the final model structure according to the significance of the variables and the Akaike information criterion (AIC). The goodness of fit was assessed for the model based on the significance test.

Results

Identification of key indicators by RDA

Table 2 shows the eigenvalues and explained variance according to RDA. As shown in Table 2, the eigenvalues for the four axes were 0.7544, 0.0101, 0.0072, and 0.0018.

Statistic	Axis 1	Axis 2	Axis 3	Axis 4
Eigenvalues	0.7544	0.0101	0.0072	0.0018
Explained variation (cumulative)	75.44	76.45	77.17	77.35
Pseudo-canonical correlation	0.8819	0.822	0.8504	0.7826
Explained fitted variation (cumulative)	97.47	98.78	99.71	99.94

Table 2. Eigenvalues and explained variance according to RDA for ecosystem degradation indicators in BYD Lake.

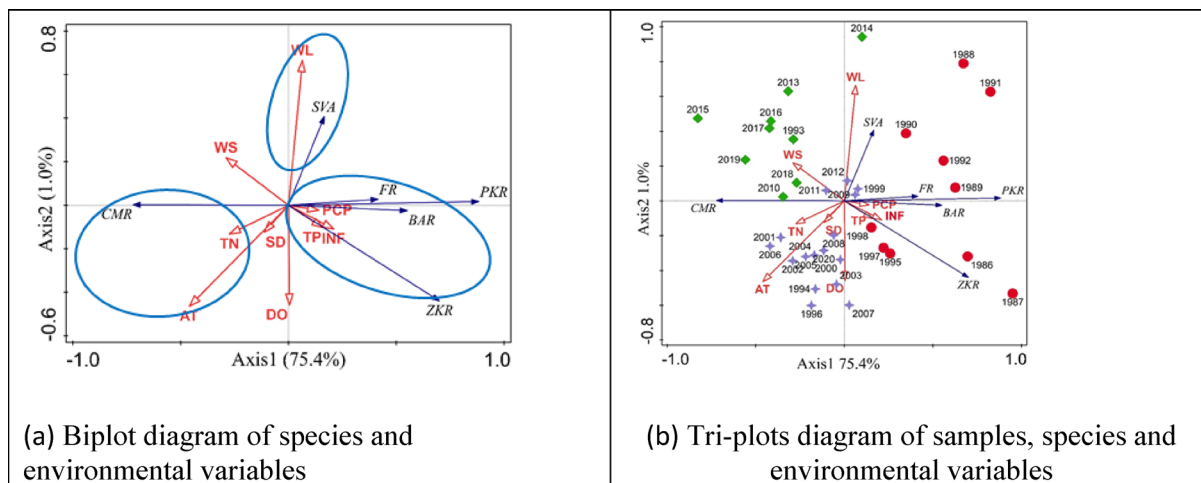


Fig. 4. The RDA analysis results of BYD Lake.

These values indicated a decrease in capacity of the RDA axes to explain the ecological response data. The explained variation represented the cumulative interpretation capacity of the ecological response data. The cumulative explained variance for the first two axes accounted for 76.45%, which means that the original ecological response data could explain 76.45% of the information. The other 23.55% of the information was attributed to other factors not considered in this study. The cumulative explained variance for the first two axes together accounted for 98.78% of the relationships between the ecological response data and explanatory variables, thereby indicating that the data were fitted well by RDA.

The relationships between species richness (blue arrows) and the driving variables (red arrows) are shown in Fig. 4a. The lower right quadrant shows that the fish richness, phytoplankton richness, benthic animal richness, and zooplankton richness were grouped together with PCP, TP, and INF, thereby suggesting that the richness of most species was strongly correlated with the hydrological regime and TP concentration. WL and submerged vegetation areas were located close to Axis 2, which suggests that submerged vegetation was mainly constrained by WL. TN, AT, and *Cyclotella meneghiniana* richness were located in the lower left quadrant, and thus TN and AT strongly affected phytoplankton due to their influence on physiological and biochemical activities. The RDA results demonstrated that the ecological response of BYD Lake was significantly influenced by various factors from 1986 to 2020, including climate change, hydrological conditions, and anthropogenic pollutants.

Notes: (1) Red arrows represent environmental factors (explanatory variables). The length of the arrow represents the intensity of the impact of the environmental factor on community changes; the longer the length, the greater the impact of the environmental factor. (2) Blue arrows represent species (response variables). The angle between the blue arrow and the red arrow can represent the correlation between the species and the environmental factor. (3) Circles, diamonds, and star-shaped points represent samples from different periods (4) ZKR- zooplankton's richness; CMR- *Cyclotella meneghiniana*'s richness; PKR- phytoplankton's richness; FR- fish's richness; SVA- submerged vegetation areas; BAR- benthic animals' richness; (5) AT- air temperature; WL- water level; INF- inflow; WS- wind speed; PCP- precipitation; TN- total nitrogen; TP- total phosphorus; DO- dissolved oxygen; SD- secchi depth.

The relationships between samples from different years, species, and environmental variables are shown in Fig. 4b. All of the samples could be divided into three groups. The samples from 1986 to 1999 denoted by red dots belonged to one group related to the hydrological conditions and TP concentration. The samples moved toward the origin during this period, thereby indicating a significant decrease in species richness. By contrast, the samples from 2010 to 2020 denoted by green diamonds moved from negative values toward the origin, and thus the species richness increased during this period. The remaining samples from 2000 to 2010 were relatively concentrated near the origin, indicating an average effect for these samples and a relatively stable state. According to the sample classification results, two state changes may have occurred during the 1990s and after 2010.

Name	Explains %	Contribution %	pseudo-F	P
AT	16.3	21	6.4	0.008
TP	10.4	13.5	4.6	0.022
INF	9.3	12	4.5	0.028
SD	14.9	19.3	9.1	0.006
TN	5.2	6.7	3.4	0.052
WL	5.8	7.5	4.3	0.038
PCP	10.6	13.6	10.4	0.006
WS	2.8	3.7	3	0.098
DO	2.1	2.7	2.3	0.178

Table 3. Relative contributions to variance and significance of variables according to RDA for ecosystem degradation indicators in BYD Lake. AT, air temperature; TP, total phosphorus; INF, inflow; SD, Secchi depth; TN, total nitrogen; WL, water level; PCP, precipitation; WS, wind speed; DO, dissolved oxygen

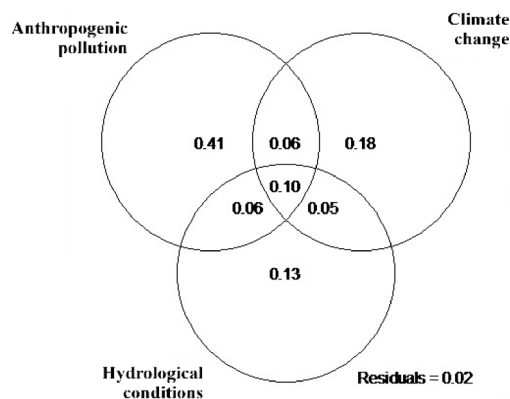


Fig. 5. The effects of anthropogenic pollution, climate change, and hydrological conditions on the ecological state of BYD Lake.

Significance tests for the variables showed that all indicators were significant at a significance level of 0.05, except for DO and WS (Table 3). The indicators with variance contributions greater than 10% were selected and sorted in descending order of variance contribution as follows: AT > SD > PCP > TP > INF. Thus, two climate change variables, two anthropogenic pollution variables, and one hydrological variable were associated with the ecological degradation of BYD Lake.

Relative contributions of variables

VPA analysis was conducted to quantify the relative contributions of the variables and the results are shown in Fig. 5.

The individual effects of anthropogenic pollution, climate change, and hydrological conditions explained 41%, 18%, and 13% of the variance, respectively. The interactions between anthropogenic pollution and climate change, anthropogenic pollution and hydrology, and hydrology and climate change also explained 6%, 6%, and 5% of the total variance, respectively. The combined effects of the three significant variables explained 10% of the total variance. The effects were ranked in the following order: anthropogenic pollution > interactions > climate change > hydrology. Thus, the ecological degradation of BYD Lake was associated with several factors, where anthropogenic pollution played a leading role and interactions ranked second.

Patterns of species richness responses to variables

The GAM method targets a single response variable, so the different types of species richness were first integrated in the CEF with a linear representation of the principal components. To assess the suitability of PCA for reducing the dimensionality of environmental variables, a Kaiser-Meyer-Olkin (KMO) test was conducted. The KMO statistic evaluates the proportion of variance shared among variables, with values closer to 1 indicating stronger collinearity and better suitability for PCA. The KMO test value of 0.672 indicated a good correlation between the variables and their suitability to the PCA method. The PCA results for the response variables are shown in Table 4.

The first two principal components retained with eigenvalues greater than 1 explained 79.4% of the total variance, as shown in Fig. 6.

The CEF represents the holistic ecological state of the ecosystem in BYD Lake. The ecosystem underwent the most significant decline from 1986 to 1995, and then remained in a relatively stable state of degradation from

Principal component	Eigenvalue	variance percentage	Cumulative variance percentage	Reserved principal component
1	3.333	55.544	55.544	3.333
2	1.429	23.816	79.361	1.429
3	0.569	9.483	88.844	
4	0.321	5.343	94.187	
5	0.249	4.157	98.345	
6	0.099	1.655	100.000	

Table 4. Total variance explained for response variables according to PCA for ecosystem degradation indicators in BYD Lake.

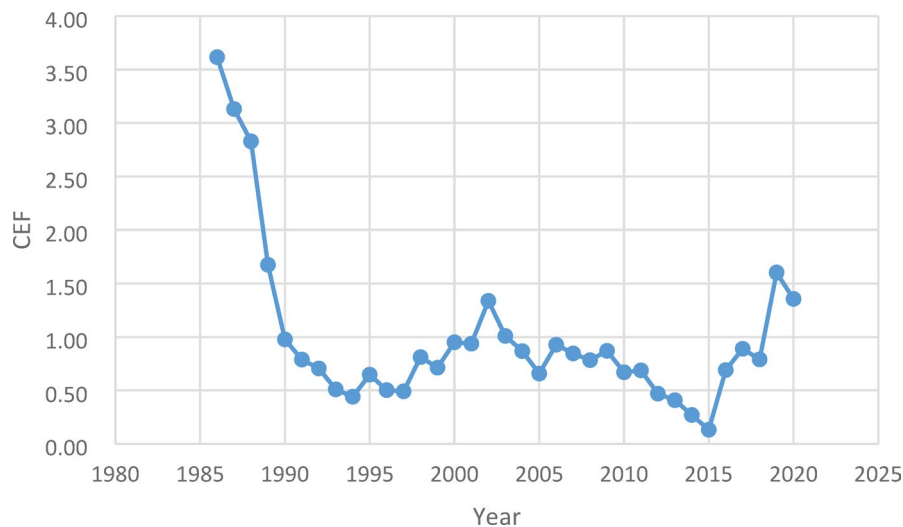


Fig. 6. The change of comprehensive evaluation function(CEF) of BYD Lake from 1986 to 2020/ (1) The time series of all species indicators can be found in the Excel file of the supplementary data. (2) CEF (Comprehensive Evaluation Function) represents the integrated result combining all species data indices. It reflects the synthesized ecosystem attribute derived from multiple response indicators via PCA, showcasing trends over years.

1996 to 2015. The ecosystem recovered slightly after reaching its worst state in 2015. The single-variable GAM analysis results obtained between the ecological state and variables in BYD Lake from 1986 to 2020 are shown in Table 5. WL and AT had highly significant ($P < 0.05$) relationships with the CEF. The TP concentration had a significant relationship ($P < 0.1$) with the CEF. The significance of the relationships between these three factors with the CEF followed the order of: WL > AT > TP. The three significant factors had nonlinear relationships with the CEF (Fig. 7). The response curve between the ecological state and WL decreased initially and then increased. At high WL values, positive correlations were found between WL and the ecological state. The response curves obtained between the ecological state and both AT and TP tended to decrease monotonically in a nonlinear manner. Thus, these results indicate that climate warming and high TP concentrations were both detrimental to the ecosystem.

In order to determine the optimal model structures, the significant variables and interactions were added to the model using the forward selection method, as shown in Table 6.

Among the six models, Model 6 significantly reduced the AIC compared with the other models. Model 6 had the highest explanatory rate (98.7%) and lowest AIC (-19.93). The AIC decreased significantly only after adding the second-order interaction effect of WL*TP to the model, which implies that the interaction between the water quantity and quality was significant. The fitted values for the optimal model significantly matched with the training data measurements (adjusted $r^2 = 0.952$, $P < 0.0001$). The validation results demonstrated an adjusted r^2 of 0.915 and $P = 0.085 < 0.1$, as shown in Fig. 8, and thus the optimal model can be used to predict the ecological state.

Discussion

Ecological degradation in lakes can be viewed as a reverse succession process in an ecosystem, with decreases in species diversity and system instability due to various types of human and natural interference. In this study, we considered the effects of driving forces consisting of hydrological conditions, anthropogenic pollution, and climate change on the ecosystem in BYD Lake while assuming that the impacts of land use patterns, habitat

Explained variable	Explanatory variable	Degrees of freedom	Explained deviation	Adj.R ²
CEF	INF	1	2.03%	-0.009
	WL	1.979	67.4%	0.654**
	TN	1.653	7.8%	0.031
	TP	1.307	14%	0.105*
	DO	1.534	6.13%	0.017
	SD	1	7.41%	0.046
	PCP	1	1.87%	-0.011
	AT	1.921	43.4%	0.4**
	WS	1.712	9.17%	0.044

Table 5. Summary of GAM results between CEF and key factors for BYD Lake. * Significant difference at $P < 0.1$ ** Highly significant difference at $P < 0.05$ GAM, generalized additive model; CEF, comprehensive evaluation function; AT, air temperature; TP, total phosphorus; INF, inflow; SD, Secchi depth; TN, total nitrogen; WL, water level; PCP, precipitation; WS, wind speed; DO, dissolved oxygen

destruction, human fishing, and invasive alien species remained unchanged or changed very little during the study period.

Identification of driving forces that affected changes in species richness

Tang et al.¹⁴ suggested that nitrogen was the limiting nutrient that affected the abundance of phytoplankton according to the TN: TP ratios. In addition, the WL can directly affect the growth of submerged vegetation by influencing photosynthesis as well as indirectly by changing the physicochemical properties of sediment³⁰. Zhang et al.²¹ indicated that fluctuations in the WL and nutrient enrichment explained changes in the state of the ecosystem in BYD Lake. Moreover, precipitation and AT extremes can directly or indirectly affect the water quality by influencing biochemical reaction rates and nutrient release from sediment. Increased precipitation can influence water quality by diluting pollutants like chemical oxygen demand (COD) and total nitrogen (TN), improving water quality. However, excessive rainfall may increase the total phosphorus (TP) concentration due to the resuspension of sediments and the release of phosphorus from lakebed sediments. This complex relationship indicates that precipitation can have both positive and negative effects on water quality, depending on its intensity and the local environmental context³⁹. Our RDA results showed that climate change, hydrological conditions, and anthropogenic pollutants were the main driving forces related to ecological change in BYD Lake, as also found in previous research.

Tri-plot diagrams based on samples, species, and environmental variables obtained by RDA were used to classify the temporal changes in species richness, which were divided into three periods: 1986–1999, 2000–2015, and 2015–2020. A sharp decline in species richness occurred in the first stage, followed by fluctuations and maintenance of the declining trend in the second stage, before slight improvements in the third stage. The pattern over time suggested that two clear transitions may have occurred during the study period, with a large one in the 1990s and a smaller one in 2015. Previous studies also showed that a steady-state transformation occurred in the 1990s, mainly due to significant fluctuations in the WL and rapid increases in nutrient concentrations^{2,21,51}. However, some evidence indicates that the water quality in BYD Lake has generally improved over the past decade⁶, and the richness of most species increased after 2015²⁸. The ecological structure and functions determined based on ecological network analysis indicated noticeable improvements in 2018 compared with 2010⁵². According to Wang et al.³⁰, the local government conducted many inter-basin water diversion actions and nutrient load reduction plans in the watershed after 2010.

Anthropogenic pollution, hydrological fluctuations, and climate change were found to be related to the ecological degradation of BYD Lake. According to VPA, anthropogenic pollution was the most significant driving force related to ecological degradation, where it explained 41% of the changes, followed by interactions with 27%, climate change factors with 18%, and hydrological fluctuations with 13%. The interaction between WL and TP was significant, thereby suggesting that the effects of the hydrological conditions influenced the ecosystem through changes in water quality. A previous study based on the responses of diatom communities showed that they were significantly influenced by anthropogenic pollutants, hydrological conditions, and climate change from 1945 to 2017². Liu et al.⁵³ found that climate change had a decisive effect on the degradation of BYD Lake, where precipitation had the most significant impact by altering the hydrological characteristics of BYD Lake. Hypothesis 1 states nutrient inputs (TP, TN) drive ecological degradation via eutrophication and biodiversity suppression, while hypothesis 2 argues climate change (temperature rise) and hydrological fluctuations (water level decline) exacerbate stress by altering physicochemical conditions. The VPA analysis shows that anthropogenic pollution (represented by TP and TN) explains 41% of the total variance in ecological degradation, ranking first among all driving factors. This directly confirms the assertion in Hypothesis 1 that nutrient input is the main driving force of ecological degradation. Climate change (temperature) and hydrological conditions (water level) explain 18% and 13% of the variance respectively, and the interaction effect between them and anthropogenic pollution is significant. They aggravate ecological pressure by changing physical and chemical conditions, providing quantitative support for Hypothesis 2.

GAM identified the significant indicators as WL, AT, and TP. Slightly different results were obtained by RDA, which identified the key factors as AT, SD, PCP, TP, and INF. However, PCP, INF, and WL are closely related, and

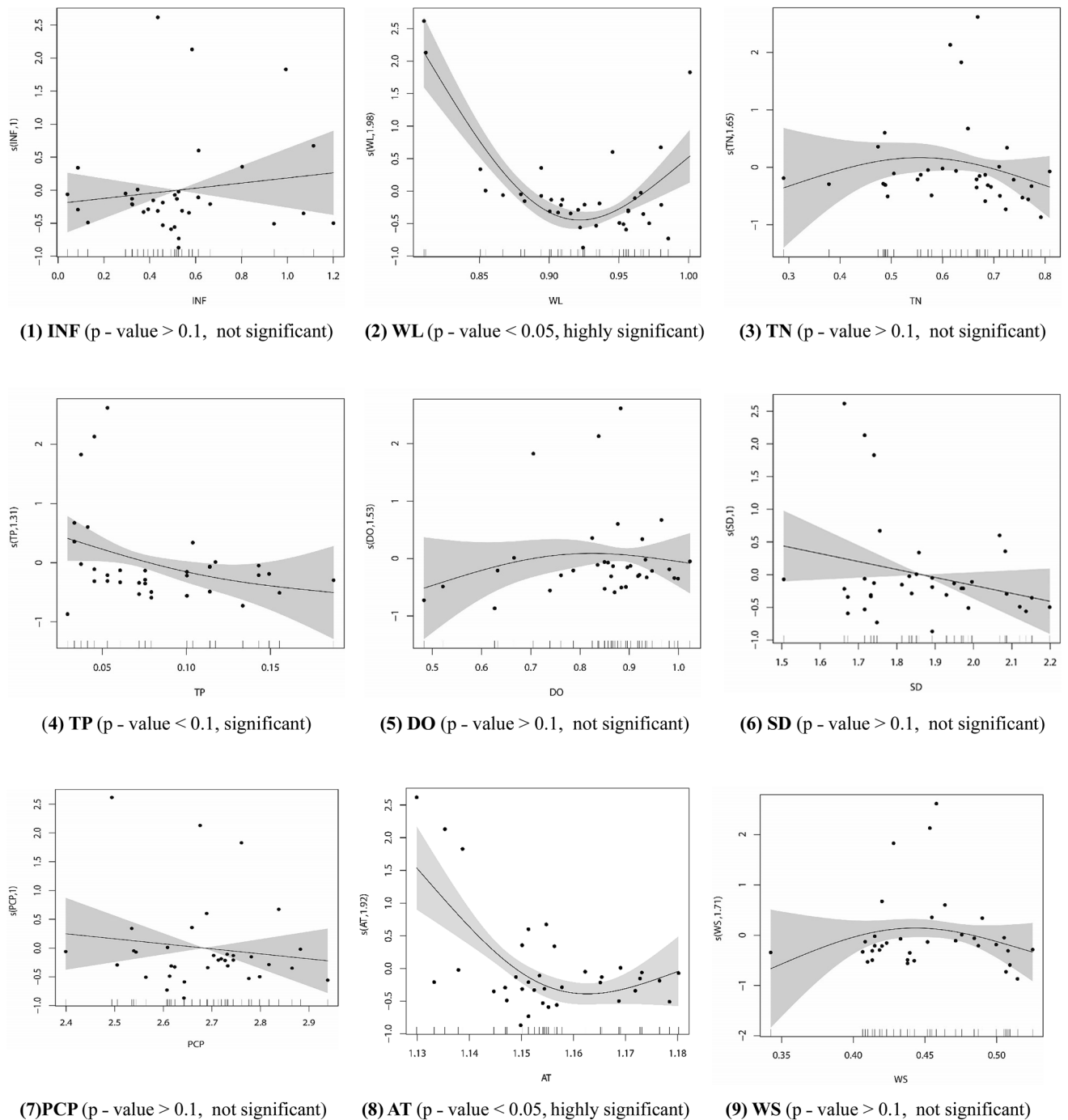


Fig. 7. The influence pattern of the single driving force to the comprehensive evaluation function (CEF). INF-inflow; WL-water level; TN- total nitrogen; TP-total phosphorus; DO-dissolved oxygen; SD-transparency; PCP-precipitation; WS-wind speed; AT- air temperature; The variables were $(\log(x+1))$ transformed before GAM.

can be approximated as the same type of variables. Therefore, the different results obtained by the two methods were only minor. These minor differences may have been due to the different response variables used by the two methods, where the former employed a synthetic variable based on the principal components and the latter used multi-dimensional species richness data. To resolve the inconsistency in significant factors identified by RDA and GAM, it is necessary to integrate the results of the two methods, focus on common significant factors such as AT and TP, and analyze them in combination with their ecological significance. Driving forces should be grouped by ecological functions (e.g., climate, hydrology, pollution) to clarify the contribution of each group and avoid interpretive biases from single factors. The dimensionality reduction approach for response variables

No.	Model structure	Deviance explained	AIC	GCV	R-sq.(adj)
1	WL	67.4%	47.99	0.2196	0.654
2	WL+TP	77.8%	37.95	0.1669	0.751
3	WL+TP+AT	79.8%	37.77	0.1688	0.761
4	WL+TP+AT+AT*TP	79.8%	37.77	0.1688	0.761
5	WL+TP+AT+WL*AT	88.4%	29.70	0.1502	0.829
6	WL+TP+AT+WL*TP	98.7%	-19.93	0.1074	0.952

Table 6. Analysis of differences in forward selection regression process results according to GAM for BYD Lake. (1) Higher explained deviance or R-sq.(adj) values indicate the better fit of a model. Smaller AIC and GCV values also indicate the better fit of a model (2) The sample size was only sufficient for one interaction term to be tested each time

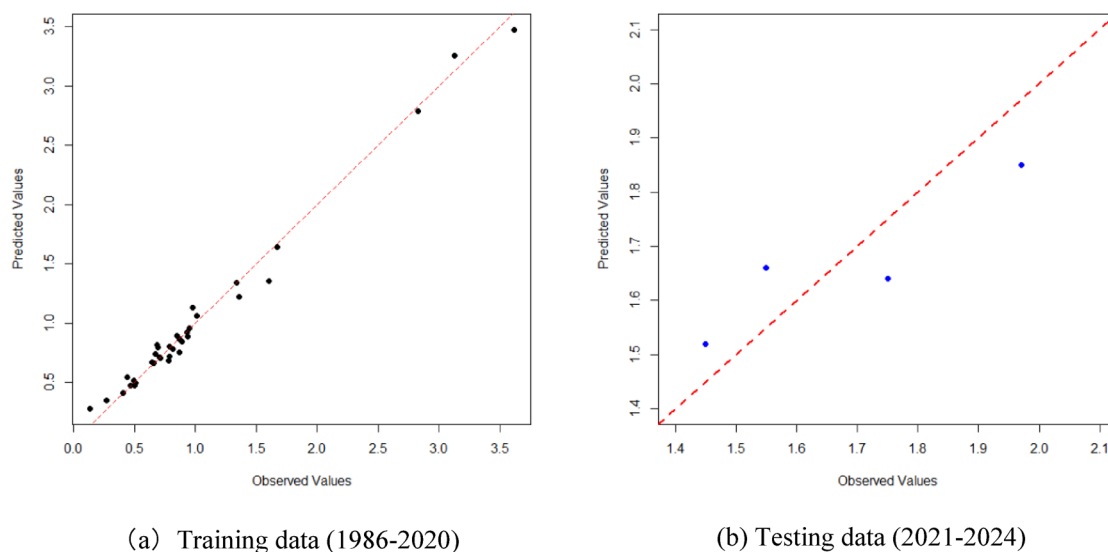


Fig. 8. Comparison of prediction of gam6 model and observed values. The observed values are CEF values calculated from the observed values of ecological responses indicators; The predicted values are CEF values predicted from GAM model. The red dashed line is $y=x$ reference line

should be optimized (such as adjusting the number of PCA principal components or highlighting key species). Model prediction errors should be compared through methods like leave-one-out cross-validation.

GAM analysis also showed that the significant factors had nonlinear effects on the ecological response. For example, when the WL was at a medium or high value, increasing the WL led to improvements in the ecological state, whereas the opposite was found when the WL was low, thereby indicating that several other factors may have affected the WL. As shown in Fig. 7, higher WL values, lower TP concentrations, and lower water temperatures corresponded to a better ecological state in BYD Lake. The variables in the models, including the WL, TP, and water temperature, and their interactions, explained the changes in the holistic ecological state of BYD Lake and can be used for making future predictions. Hypothesis 3 is examined using GAM's smoothing curves, showing species richness response to WL and AT had a stronger threshold effect than that of other variables.

Effectiveness and applicability of the proposed analytical framework

The conventional correlation analysis and multiple regression methods cannot handle nonlinear evolutionary processes with multiple inputs and outputs, and complex interactions among factors. Thus, we integrated a combination of new statistical techniques, i.e., RDA, VPA, and PCA-based GAM, to assess the causal relationships between various factors and response variables in BYD Lake. RDA and VPA were combined to identify the main factors and the relative contribution of each factor. These methods are suitable for analyzing the relationships between multiple independent variables and multiple response variables. PCA-based GAM was employed to further understand the patterns related to key variables and to optimize the structure of the prediction model to obtain the best results. PCA-based GAM is a non-parametric statistical method that can detect nonlinear relationships between variables when the functional relationships between explanatory and response variables are unclear. Our results demonstrated the effectiveness and feasibility of the proposed analytical framework. In contrast, linear methods may oversimplify ecological processes, potentially leading to incomplete interpretations and suboptimal management strategies. For example, holistic ecosystem states, interactive effects and the

threshold effects. The proposed analytical framework's ability to handle complexity makes it particularly suitable for degraded ecosystems like BYD Lake, where multiple stressors interact nonlinearly to drive ecological change.

The integration of Principal Component Analysis (PCA) and Generalized Additive Models (GAM) represents a methodological innovation in this study, offering advantages as follows: PCA compresses multi-dimensional ecological response indicators (e.g., phytoplankton, fish, benthic animal richness) into a Composite Evaluation Function (CEF) via principal components (PCs), reducing collinearity and mitigating overfitting risks in GAM due to high-dimensional data. GAM quantifies nonlinear relationships between CEF and drivers by using smoothing functions, overcoming assumptions of traditional linear models. For example, WL exhibits a threshold effect (decline followed by increase) on CEF, while AT and TP show monotonic nonlinear declines—findings unrevealed by linear models. The CEF, acting as a proxy for ecosystem states, integrates responses across trophic levels, avoiding single-indicator bias. This PCA-GAM framework is not only applicable to BYD Lake but also generalizable to other lakes under multi-source drivers, providing a methodological reference for shallow lake management in the context of global change.

In this study, we assumed that the combined species richness could be used as an effective description of the holistic ecological state of BYD Lake. Characterizing the ecological state of lakes under management practices remains challenging and long-term biological data are often difficult to obtain. Moreover, other holistic indicators, such as ecological structure and function indicators, ecological network analysis indicators, and diversity and stability indicators, are often not available due to the requirement for large amounts of information about species distributions and interspecies interactions. Therefore, the use of multi-source data is encouraged. New data acquisition methods, such as remote sensing and sediment core analysis, can be employed to obtain long-term biological information. Machine learning methods for interpolation can also be applied to complete time series with missing data⁵⁴. In addition, the BYD Lake, as a semi-arid shallow lake shaped by upstream reservoir interception and intensive agriculture, introduces some limitations. First, its regional specificity may restrict generalizability to other lake types (e.g., deep or oligotrophic systems), as drivers like water level-total phosphorus interactions could behave differently in hydrologically connected ecosystems. Second, while the 35-year dataset captures decadal trends, multi-century paleolimnological records are needed to fully characterize long-term climate impacts. Third, some excluded drivers (e.g., land use change, invasive species) potentially confounded conclusions. Future research should validate the RDA-VPA-GAM framework in contrasting other lakes, extend temporal coverage to account for unmeasured variables under global change scenarios. Based on the main factors and response patterns identified in this study, hydrological management, water quality restoration, and biological manipulation are recommended for the management of BYD Lake. The minimum ecological water level of BYD Lake is 7.45 ± 0.66 m, the suitable water level is 8.61 ± 0.52 m, and the maximum water level is 9.46 ± 0.51 m⁵⁵. Our research results show that species richness increases when the water level is from 6.9 m to 9.0 m. A suitable WL can be maintained through various means, such as water transfer and reservoir operation. Hydrological management can directly improve the ecological state but priority should be given to reducing the nutrient loadings rather than other practices according to the GAM results. Thus, nutrient loading abatement is the most critical measure for improving the ecological state of BYD Lake. Various methods can be applied to remove nutrients from the lake watershed, such as agricultural non-point source pollution control, wetland restoration, nitrogen and phosphorus removal from domestic sewage, and dilution⁵⁶. However, the feedback between multiple factors should be comprehensively considered due to the significant interactive effects between water quality, hydrology, and climate. More research is required to understand the responses of lake ecosystems to climate change, which will help to clarify the mechanisms involved and promote ecosystem-based adaptation and mitigation. Bio-manipulation has also been proposed to constrain freshwater primary producers boosted by eutrophication^{57–59}. Moreover, understanding ecosystem changes requires long-term monitoring of ecosystems.

The Composite Evaluation Function (CEF) offers actionable solutions for lake management. It enables real-time ecological health monitoring, triggering alerts (e.g., $CEF < 0.5$ in 2015) for urgent interventions like nutrient reduction. The CEF supports policy by classifying lakes into health tiers (e.g., $CEF > 0.8$ = healthy) for targeted resource allocation and predicting climate impacts (e.g., CEF drops 12% per 1 °C temperature rise). Adaptive strategies include TP control and optimal WL management (WL < 6.9 m threshold triggering degradation). Unlike single-metric approaches, CEF integrates multi-trophic responses, revealing hidden declines. For example, during 2006–2015, although chlorophyll-a (Chl-a) concentrations in BYD Lake decreased, the Composite Evaluation Function (CEF) still indicated ecological degradation (Fig. 9).

Conclusion

In this study, we analyzed the effects of driving forces related to changes in the species richness in BYD Lake during the 35 years from 1986 to 2020. An integrated analytical framework was developed to screen, identify, and predict the driving forces related to degradation of the ecosystem in the lake. RDA and VPA were employed to determine the key factors and their relative contributions to explaining the total variance in the species richness. PCA-based GAM was applied to explore the patterns associated with the effects of key factors on species richness. The results showed that submerged vegetation was mainly constrained by WL. TN and AT strongly impacted phytoplankton, and the hydrological regime and TP strongly affected the richness of other species. RDA showed that anthropogenic pollution, climate change, and hydrological conditions significantly influenced the communities in BYD Lake. VPA demonstrated that anthropogenic pollution, climate change, and hydrological conditions explained 41%, 18%, and 13% of the total variance, respectively. GAM showed that the most significant factors were WL, AT, and TP, and they affected the changes in the ecological state of the lake in a nonlinear manner. The best predictive equation containing WL, AT, and TP as the three main factors, and the interaction term of WL*TP explained 98.4% of the total variance. A higher WL, lower TP, and lower AT corresponded to better ecological conditions. Thus, priority should be given to improving the water quality when selecting adaptive management measures. Future research should combine modeling with the existing

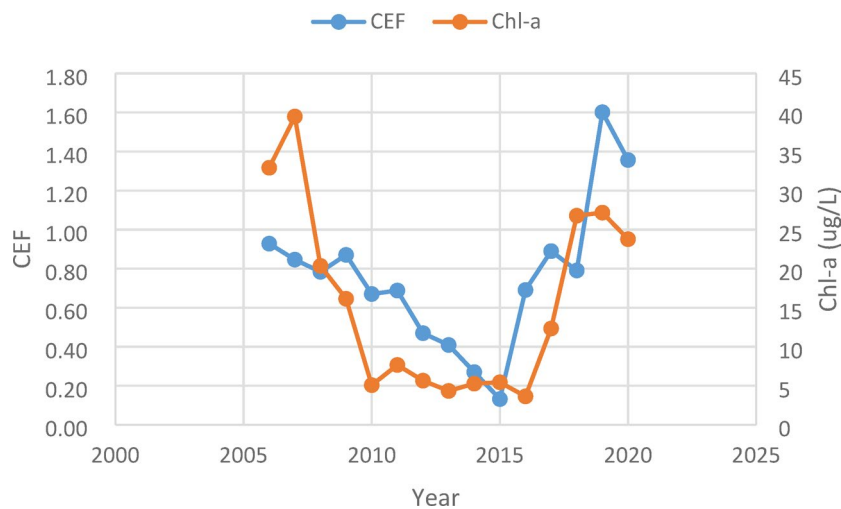


Fig. 9. The time series of annual mean values of chlorophyll a and CEF (2005–2020). CEF - a comprehensive evaluation function for ecological response indicators; Chl-a - chlorophyll a.

mechanistic analysis methods to confirm the reliability of the results obtained and extend temporal coverage to account for unmeasured variables under global change scenarios.

Data availability

Data is provided within the supplementary information files.

Received: 10 December 2024; Accepted: 29 July 2025

Published online: 25 January 2026

References

- Wang, X. et al. Water quality variation and driving factors quantitatively evaluation of urban lakes during quick socioeconomic development. *J. Environ. Manage.* **344**, 118615 (2023).
- Mao, X. et al. Abrupt diatom assemblage shifts in Lake Baiyangdian driven by distinct hydrological changes and yet more so by gradual eutrophication. *Limnologia* **105**, 126155 (2024).
- Baker, M. E. & King, R. S. A new method for detecting and interpreting biodiversity and ecological community thresholds. *Methods Ecol. Evol.* **1**, 25–37 (2010).
- Kang, J. et al. How do natural and human factors influence ecosystem services changing? A case study in two most developed regions of China. *Ecol. Indic.* **146**, 109891 (2023).
- Tang, C., Yi, Y., Yang, Z., Zhang, S. & Liu, H. Effects of ecological flow release patterns on water quality and ecological restoration of a large shallow lake. *J. Clean. Prod.* **174**, 577–590 (2018).
- Han, Q. et al. Anthropogenic influences on the water quality of the Baiyangdian lake in North China over the last decade. *Sci. Total Environ.* **701**, 134929 (2020).
- Han, Q. et al. Assessing alterations of water level due to environmental water allocation at multiple Temporal scales and its impact on water quality in Baiyangdian Lake, China. *Environ. Res.* **212**, 113366 (2022).
- Liu, L. & You, X. Water quality assessment and contribution rates of main pollution sources in Baiyangdian Lake, Northern China. *Environ. Impact Assess. Rev.* **98**, 106965 (2023).
- Wei, Z., Yu, Y. & Yi, Y. Analysis of future nitrogen and phosphorus loading in watershed and the risk of lake blooms under the influence of complex factors: implications for management. *J. Environ. Manage.* **345**, 118662 (2023).
- Liu, C., Liu, L. & Shen, H. Seasonal variations of phytoplankton community structure in relation to physico-chemical factors in lake Baiyangdian, China. *Procedia Environ. Sci.* **2**, 1622–1631 (2010).
- Han, Z. & Cui, B. Performance of macrophyte indicators to eutrophication pressure in ponds. *Ecol. Eng.* **96**, 8–19 (2016).
- Yang, W., Yan, J., Wang, Y., Zhang, B. T. & Wang, H. Seasonal variation of aquatic macrophytes and its relationship with environmental factors in Baiyangdian Lake, China. *Sci. Total Environ.* **708**, 135112 (2020).
- Sun, B. et al. Integrated modeling framework to evaluate the impacts of multi-source water replenishment on lacustrine phytoplankton communities. *J. Hydrol.* **612**, 128272 (2022).
- Tang, C. et al. Planktonic indicators of trophic States for a shallow lake (Baiyangdian lake, China). *Limnologia* **78**, 125712 (2019).
- Yan, S. et al. A hybrid PCA-GAM model for investigating the Spatiotemporal impacts of water level fluctuations on the diversity of benthic macroinvertebrates in Baiyangdian Lake, North China. *Ecol. Indic.* **116**, 106459 (2020).
- Yang, Y. et al. Spatio-temporal variations of benthic macroinvertebrates and the driving environmental variables in a shallow lake. *Ecol. Indic.* **110**, 105948 (2020b).
- Liao, Z. et al. An integrated simulation framework for NDVI pattern variations with dual society-nature drives: A case study in Baiyangdian Wetland, North China. *Ecol. Indic.* **158**, 111584 (2024).
- Yang, W., Yang, Z. & Qin, Y. An optimization approach for sustainable release of e-flows for lake restoration and preservation: model development and a case study of Baiyangdian lake, China. *Ecol. Modell.* **222**, 2448–2455 (2011).
- Yang, W. & Yang, Z. Effects of long-term environmental flow releases on the restoration and preservation of Baiyangdian lake, a regulated Chinese freshwater lake. *Hydrobiologia* **730** (1), 79–91 (2014).
- Liu, X., Yang, W., Fu, X. & Li, X. Determination of the ecological water levels in shallow lakes based on regime shifts: A case study of china's Baiyangdian lake. *Ecohydrol Hydrobiol.* **24** (4), 931–943. <https://doi.org/10.1016/j.ecohyd.2023.08.014> (2024).
- Zhang, X., Yi, Y. & Yang, Z. The long-term changes in food web structure and ecosystem functioning of a shallow lake: implications for the lake management. *J. Environ. Manage.* **301**, 113804 (2022).

22. Zhang, Q., Zhang, Y., Yu, T. & Zhong, T. Primary driving factors of ecological environment system change based on directed weighted network illustrating with the Three-River headwaters region. *Sci. Total Environ.* **916**, 170055 (2024).
23. Zhang, X., Yi, Y., Yang, Y., Liu, H. & Yang, Z. Modelling phosphorus loading to the largest shallow lake in Northern China in different shared socioeconomic pathways. *J. Clean. Prod.* **297**, 126537 (2021).
24. Cai, Y. et al. How does water diversion affect land use change and ecosystem service: A case study of Baiyangdian wetland, China. *J. Environ. Manage.* **344**, 118558 (2023).
25. Yang, Y., Yin, X. & Yang, Z. Environmental flow management strategies based on the integration of water quantity and quality, a case study of the Baiyangdian Wetland, China. *Ecol. Eng.* **96**, 150–161 (2016).
26. Hu, S., Wang, X. & Song, X. Could the hydrological conditions of lake Baiyangdian support a booming metropolis? *Sci. Total Environ.* **869**, 161764 (2023).
27. Yang, W., Tian, Y., Zhang, Z., Liu, Q. & Zhao, Y. Evolution of phytoplankton community and biotic integrity in Baiyangdian lake in recent 60 years. *Environ. Ecol.* **1**(8), 1–9 (2019).
28. Yi, Y., Lin, C. & Tang, C. Hydrology, environment and ecological evolution of lake Baiyangdian since 1960s. *J. Lake Sci.* **32**(5), 1333–1347 (2020).
29. Zeng, Y., Zhao, Y. & Qi, Z. Evaluating the ecological state of Chinese lake Baiyangdian (BYD) based on ecological network analysis. *Ecol. Indic.* **127**, 107788 (2021).
30. Wang, Y., Gong, Z. & Zhou, H. Long-term monitoring and phenological analysis of submerged aquatic vegetation in a shallow lake using time-series imagery. *Ecol. Indic.* **154**, 110646 (2023).
31. Wang, Y. et al. Fish community structure and its relationship with environmental factors in Baiyangdian lake. *J. Shanghai Ocean. Univ.* **31**(6), 1488–1501 (2022).
32. Šmilauer, P. & Lepš, J. *Multivariate Analysis of Ecological Data Using CANOCO 5* 2nd edn (Cambridge University Press, 2014).
33. Fan, K. et al. Soil biodiversity supports the delivery of multiple ecosystem functions in urban greenspaces. *Nat. Ecol. Evol.* **7**, 113–126 (2023).
34. Wood, S. N. *Generalized Additive Models: an Introduction with R, Second Edition* 2nd edn (Chapman and Hall/CRC, 2017).
35. Guisan, A., Edwards, T. C. Jr & Hastie, T. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol. Modell.* **157**, 89–100 (2002).
36. Xu, M. et al. The ecological degradation and restoration of Baiyangdian Lake, China. *J. Freshw. Ecol.* **13** (4), 433–446 (1998).
37. Zhang, X., Yi, Y., Cao, Y. & Yang, Z. Disentangling the effects of phosphorus loading on food web stability in a large shallow lake. *J. Environ. Manage.* **328**, 116991 (2023).
38. Yang, J. et al. What is the pollution limit? Comparing nutrient loads with thresholds to improve water quality in lake Baiyangdian. *Sci. Total Environ.* **807**, 150710 (2022).
39. Han, Y. & Bu, H. The impact of climate change on the water quality of Baiyangdian lake (China) in the past 30 years (1991–2020). *Sci. Total Environ.* **870**, 161957 (2023).
40. Vatcheva, K. P., Lee, M., McCormick, J. B. & Rahbar, M. H. Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology* **6** (2), 227. <https://doi.org/10.4172/2161-1165.1000227> (2016).
41. Oshun, M. I. & Grantham, T. E. Leveraging species richness and ecological condition indices to guide systematic conservation planning. *J. Environ. Manage.* **341**, 117970 (2023).
42. Myllyviita, T. et al. Assessing biodiversity impacts in life cycle assessment framework -Comparing approaches based on species richness and ecosystem indicators in the case of Finnish boreal forests. *J. Clean. Prod.* **236**, 117641 (2019).
43. Niu, Y., Schuchardt, M. A., Heßberg, A. & Jentsch, A. Stable plant community biomass production despite species richness collapse under simulated extreme climate in the European alps. *Sci. Total Environ.* **864**, 161166 (2023).
44. Jan, L. *Multivariate Analysis of Ecological Data Using CANOCO 50–51* (Cambridge University Press, 2003).
45. Gabarron, M. & Acosta, A. F. J. A. Use of multivariable and redundancy analysis to assess the behavior of metals and arsenic in urban soil and road dust affected by metallic mining as a base for risk assessment. *J. Environ. Manage.* **206**, 192–201 (2018).
46. Wang, J., Wang, H., Cao, Y., Bai, Z. & Qin, Q. Effects of soil and topographic factors on vegetation restoration in opencast coal mine dumps located in a loess area. *Sci. Rep.* **6**, 22058 (2016).
47. Yi, Y., Sun, J. & Zhang, S. A habitat suitability model for Chinese sturgeon determined using the generalized additive method. *J. Hydrol.* **534**, 11–18 (2016).
48. Murphy, R. R., Perry, E., Harcum, J. & Keisman, J. A generalized additive model approach to evaluating water quality: Chesapeake Bay case study. *Environ. Modell. Softw.* **118**, 1–13 (2019).
49. Tasnim, R. et al. Site suitability mapping for different seaweed cultivation systems along the coastal and marine waters of bangladesh: A generalized additive modelling approach for prediction. *Algal Res.* **78**, 103404 (2024).
50. Zhao, Y. et al. Modeling and application of sensory evaluation of blueberry wine based on principal component analysis. *Curr. Res. Food Sci.* **6**, 100403 (2023).
51. Yang, Y., Yin, X., Yang, Z., Sun, T. & Xu, C. Detection of regime shifts in a shallow lake ecosystem based on multi-proxy paleolimnological indicators. *Ecol. Indic.* **92**, 312–321 (2018).
52. Guo, S. Y., Wang, J. G., Wang, Y., Chen, Z. & Yan, J. Analysis of the ecosystem structure and energy flow of the Baiyangdian lake in recent 10 years based on the ecopath model. *Asian J. Ecotoxicol.* **15**(5), 169–180 (2020).
53. Liu, C., Xie, G. & Xiao, Y. Impact of climatic change on Baiyangdian wetland. *Resour. Environ. Yangtze Basin* **16**(2), 245–250 (2007).
54. Lepot, M., Aubin, J. B. & Clemens, F. H. L. R. Interpolation in time series: an introductory overview of existing Methods, their performance criteria and uncertainty assessment. *Water* **9** (10), 796 (2017).
55. Yang, W., Zhao, Y., Liu, Q. & Sun, T. A systematic literature review and perspective on water-demand for ecology of lake Baiyangdian. *J. Lake Sci.* **32**(2), 294–308. <https://doi.org/10.18307/2020.0202> (2020).
56. Yang, J., Stokal, M., Kroeze, C., Bai, Z. & Ma, L. Nutrient and manure management to improve water quality in urbanizing Baiyangdian. *Nutr. Cycl. Agroecosyst.* **127**, 51–67 (2023).
57. Shapiro, J., Lamarra, V. & Lynch, M. Biomanipulation: an ecosystem approach to Lake restoration. In: Brezonik PL, Fox JL (eds) Proceedings of a symposium on water quality management through biological control. *University of Florida, Gainesville*, pp 85–96. (1975).
58. Triest, L., Stiers, I. & Onsem, S. V. Biomanipulation as a nature-based solution to reduce cyanobacterial blooms. *Aquat. Ecol.* **50**, 461–483 (2016).
59. Peng, G. et al. Ecosystem stability and water quality improvement in a eutrophic shallow lake via long-term integrated biomanipulation in Southeast China. *Ecol. Eng.* **159**, 106119 (2021).

Acknowledgements

We thank International Science Editing (<http://www.internationalscienceediting.com>) for language editing this manuscript.

Author contributions

Yong Zeng: Writing - original draft, Methodology, Writing - Review & Editing; Yanwei Zhao: Conceptualization,

Investigation, Data Curation; Wei Yang: Formal analysis, Software; All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-026-37179-6>.

Correspondence and requests for materials should be addressed to Y.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026