Science of the Total Environment 912 (2024) 168717

Contents lists available at ScienceDirect

Science of the Total Environment

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

Spatially explicit estimation of freshwater fish stock biomass with limited data: A case study of an endangered endemic fish on the Tibetan Plateau, China

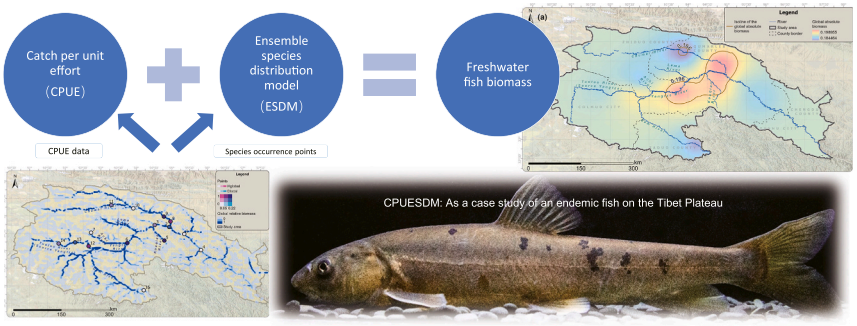
Kunyuan Wanghe [a,*,1], Shahid Ahmad [b,g,1], Xin Zhou [e,1], Fei Tian [a], Sijia Liu [a], Bingzheng Zhou [e], Ghulam Nabi [c], Guojie Wang [d], Kemao Li [d], Shenglong Jian [d], Huamin Jiang [f], Shengxue Chen [a], Yimeng Niu [a], Muhammad Ismail Khan [h], Kai Zhao [a,*]

[a] Key Laboratory of Adaptation and Evolution of Plateau Biota, Laboratory of Plateau Fish Evolutionary and Functional Genomics, Qinghai Key Laboratory of Animal Ecological Genomics, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, China
[b] School of Ecology and Environment, Hainan University, Haikou, China
[c] Institute of Nature Conservation, Polish Academy of Sciences, Krakow, Poland
[d] Qinghai Provincial Fishery Technology Extension Center, Xining, China
[e] Qinghai University, Xining, China
[f] The Rescues Center of Qinghai-Lake Naked Carp, Xining, China
[g] Wildlife and Ecosystem Research Lab, Department of Zoology, University of Chitral, Khyber Pakhtunkhwa, Pakistan
[h] Department of Zoology, Islamia College, Peshawar, Pakistan

HIGHLIGHTS

- We introduce integrated modeling to estimate fish biomass using limited data.
- We estimated *Herzensteinia microcephalus* biomass and spatial traits in a case study.
- The validation results showed high accuracy between observed and simulated values.
- Our method offers key insights for effective fishery and conservation management.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Fernando A.L. Pacheco

Original content: Data for: Spatial explicit estimation of endangered freshwater fish biomass with limited data, as a case study of an

ABSTRACT

Accurate evaluation of fish stock biomass is essential for effective conservation management and targeted species enhancement efforts. However, this remains challenging owing to limited data availability. Therefore, we present an integrated modeling framework combining catch per unit effort with ensemble species distribution modeling called CPUESDM, which explicitly assesses and validates the spatial distribution of stock biomass for freshwater

Abbreviations: CPUE, Catch per unit effort; ESDM, Ensemble species distribution modeling; CPUESDM, Catch per unit effort with ensemble species distribution modeling; eDNA, Environmental DNA; ROC, Receiver operating characteristic; SRTM, Shuttle radar topography mission.

* Corresponding authors.
E-mail addresses: wanghekunyuan@gmail.com, wanghekunyuan@189.cn (K. Wanghe), zhaokai@nwipb.cas.cn (K. Zhao).

[1] The authors contributed equally as the first authors.

https://doi.org/10.1016/j.scitotenv.2023.168717
Received 6 September 2023; Received in revised form 17 November 2023; Accepted 18 November 2023
Available online 25 November 2023

endemic fish on the Tibet Plateau, China (Original data)

Keywords:
Biomass estimation
Catch per unit effort
Freshwater fish conservation
Herzensteinia microcephalus
Species distribution model
Yangtze River Basin

fish species with limited data, applied to *Herzensteinia microcephalus*. The core algorithm incorporates the Leslie regression model, ensemble species distribution modeling, and exploratory spatial interpolation techniques. We found that *H. microcephalus* biomass in the Yangtze River source area yielded an initial estimate of 113.52 tons. Our validation results demonstrate high accuracy with a Cohen's kappa coefficient of 0.78 and root mean square error of 0.05. Furthermore, our spatially-explicit, global, absolute biomass density map effectively identified areas with high and low concentrations of biomass distribution centers. Additionally, this study offers access to the source code, example raw data, and a step-by-step instruction manual for other researchers using field data to explore the application of this model. Our findings can help inform for future conservation efforts around fish stock biomass estimation, especially for endangered species.

1. Introduction

Evaluating fish stock biomass is pivotal for effective conservation management and targeted species enhancement (Li et al., 2020; Radinger et al., 2023). For example, national long-term fishing monitoring data from the Yangtze River showed the current fish stock biomass in the Yangtze Basin is approximately 124,800 tons, representing only 27.3 % of its historical peak in the 1950s (Huang and Li, 2016; Zhang et al., 2020). These findings emphasize the urgent need to implement proactive measures to safeguard our aquatic ecosystems and have promoted certain fishery management policies/projects (Zhang et al., 2023), such as the ten-year fishing moratorium in the Yangtze Basin effective as of January 1, 2020, issued by Ministry of Agricultural and Rural Affairs of China (Wang et al., 2022).

The study area is the source of the Yangtze River, encompassing its headstream and three tributaries (the Tuotuo, Chumar, and Dam rivers) located in the northeastern part of the Tibetan Plateau, Qinghai Province, China. Recent studies have reported that 69.0 % of the threatened and extinct fish species are endemic to the Yangtze Basin (Huang and Li, 2016). Within this, the proportion of threatened fish species in the Yangtze Basin is 21.1 %, encompassing 107 endangered species. In contrast, the proportion of endangered fish species is as high as 50.0 % in the source region of the Yangtze (Liang et al., 2016). However, several studies (Chen et al., 2022; Kindong et al., 2020; Yue et al., 2021) have overlooked the assessment of fish stock biomass in the source region of the Yangtze. Owing to inherent natural and logistical limitations, current methodologies have failed to provide sufficient support for accurate stock biomass estimation in this highland region. It is widely acknowledged as a crucial region for conserving freshwater biodiversity in the Yangtze River Basin. However, accurately estimating the stock biomass of the freshwater fish in the study area remains a challenge.

To date, fish stock biomass estimation has primarily relied on high-frequency fish-catching data (Mahadevan et al., 2019), environmental DNA (eDNA) (Doi et al., 2015), and acoustic fish stock assessment (Block et al., 2019). First, high-intensity invasive fishing violates the principles of conservation biology concerning endangered species, leading to physical damage or significant stress on individuals, thereby impacting their welfare and growth. Second, acoustic fish stock assessments usually require navigable waters, such as downstream rivers, lakes, and oceans. However, shipping an echosounder/sonar for fish stock biomass monitoring to upstream mountainous freshwater areas is impractical because of the steep riverbed gradients and poor navigation conditions. Third, Takahara et al. (2012) initially proposed the use of eDNA technology to assess fish stock biomass. This approach assumes a direct relationship between aquatic vertebrate biomass and the release of eDNA into the water at a specific rate (Li et al., 2020). Consequently, measuring eDNA concentrations only provides information on relative fish abundance rather than absolute biomass (Murakami et al., 2019). In conclusion, the limitations imposed by sampling conditions complicate the estimation of the stock biomass of freshwater fish. For example, all available records of *Herzensteinia microcephalus* consist of dozens of occurrence points scattered across the Tibetan Plateau (Zhu et al., 2021), an endangered Cyprinid fish listed in the Red List of China's Vertebrates (Jiang et al., 2016). Therefore, it is imperative to develop a rapid and cost-effective approach for assessing the stock biomass of freshwater fish with limited data, especially for endangered fish species.

Ensemble species distribution modeling (ESDM), which is based on machine-learning algorithms, offers a feasible solution to this issue (Ahmad et al., 2020). ESDM effectively models the global habitat suitability with limited occurrence data based on non-intrusive fishing monitoring records (Tikhonov et al., 2020). Several studies (Hubert and Rahel, 1989; Kelly et al., 2015; Stewart et al., 2005) have used ESDM, confirming a positive correlation between the biomass/abundance and habitat suitability of aquatic organisms; therefore, habitat suitability output by ESDM robustly represents global relative biomass. Additionally, the catch per unit effort (CPUE) is a commonly used measure of abundance to support fishery management (Říha et al., 2023). CPUE data have demonstrated a robust correlation with fish stock biomass (Emmrich et al., 2012), enabling the estimation of the local absolute biomass within a closed system through repeated sampling over consecutive short time intervals (Arreguín-Sánchez, 1996; Leslie and Davis, 1939). These two approaches combined then establishing a bridge between the global relative biomass and local absolute biomass.

Therefore, this study presents a novel integrated modeling framework, namely, combining catch per unit effort with ensemble species distribution modeling (CPUESDM), to estimate the stock biomass of freshwater fish species cost-effectively and spatially explicitly with limited data. We share the source code and example input data as a case study focusing on *H. microcephalus*, an endangered freshwater fish species endemic to the Tibetan Plateau.

Several endangered and endemic fish species, including *Acipenser dabryanus*, *Hucho bleekeri*, and *H. microcephalus*, are fish stocking[2] in the Yangtze River Basin. However, despite the existence of fish stock biomass estimation data, based on fishery catch records, since the 1950s across various sections of the Yangtze River Basin (Huang and Li, 2016; Zhang et al., 2020), fish stock biomass in our specific study area has not previously been estimated because of limited access to fishery catch information (Fig. 2) (Dong et al., 2023). Therefore, here, we present a case study estimating the biomass of *H. microcephalus* to support the fish stocking and contribute to the ten-year fishing moratorium in the Yangtze Basin. In summary, this study aimed 1) to propose an integrated modeling framework named CPUESDM for accurately estimating the freshwater fish stock biomass with limited data, and 2) to identify and calibrate the stock biomass and spatial distribution priority of *H. microcephalus*.

2. Materials and methods

2.1. Aim, development, and usage of CPUESDM

This study introduces a novel modeling framework, CPUESDM, aimed at the spatially explicit estimation and validation of the stock biomass of freshwater fish with limited data. The required inputs included species occurrence points (Table S2 in Supplementary material

[2] The term fish stocking is the practice of releasing fish that are artificially raised into a natural body of water, to supplement existing wild populations.

1) and CPUE data (Table S4 and Fig. S3 in Supplementary material 1). The CPUESDM script (available in the Section of Data availability) was developed using Python 3.9.16 on ArcGIS Pro (Esri, Redlands, California, USA). A comprehensive step-by-step instruction manual is provided in the Supplementary material 2 outlining the replication of CPUESDM using user fieldwork data. By adapting the input and corresponding parameters in the script to their specific datasets, users can readily estimate the stock biomass of other freshwater fish species. Alternatively, if users lack expertise in scripting Python code, they may utilize the built-in toolbox within the ArcGIS Pro software to conduct these analyses, as outlined in the step-by-step instruction manual.

2.2. Modeling framework of CPUESDM

CPUESDM is an integrated model, combining the CPUE and ESDM approaches (Fig. 1), based on a core algorithm that includes the Leslie regression model (Leslie and Davis, 1939) (Sections 2.4 and 2.5), ESDM (Ahmad et al., 2020) (Section 2.6), and exploratory spatial interpolation techniques (Section 2.8) (Alam et al., 2016; Yin et al., 2022). These two steps were conducted independently; thus, the CPUESDM employed a generalized linear model to cross-check and cross-calibrate the two independent calculations (Section 2.7). Subsequently, the calibrated local absolute biomass was translated into the global absolute biomass using the optimal solution from the 20 spatial interpolation methods (Section 2.8). Finally, map algebra was used to extract the total biomass of the designated water region (Section 2.9). By integrating these commonly used individual models for estimating aquatic organism biomass and their spatial patterns, CPUESDM ensures robustness in the modeling process while leveraging their combined strengths to improve overall modeling performance.

2.3. Study area

The study area, located in the northeastern part of the Tibetan Plateau, Qinghai Province, China, covers an area of 157,718 km^2, with elevations ranging from 3399 to 6350 m, and serves as the source of the Yangtze River Basin (Fig. 2). The Yangtze River is the longest river in Eurasia and the third longest river worldwide. The Yangtze River originates from within the study area, along with its headwaters and several tributaries. Our previous study (Feng et al., 2023; Tang et al., 2019; Zhao et al., 2009) indicated that because of convergent evolutionary processes and Pleistocene glaciation events, a distinct biogeographic pattern has emerged among the Cypriniformes fishes inhabiting this study area. Notably (Table S1 in Supplementary material 1), 45.0 % of the fish species are endemic to the Yangtze Basin, whereas 50.0 % are classified as threatened according to either the International Union for Conservation of Nature Red List or the Red List of China's Vertebrates (Jiang et al., 2016).

2.4. Fieldwork sampling and CPUE data capture

The CPUE data were collected through fieldwork sampling conducted between 2021 and 2022 using two parallel-set gillnets and four fish cages for fish catching. A schematic diagram of this sampling method is shown in Fig. S1, Supplementary material 1. The deployment of the two parallel-set gillnets created a closed waterway that effectively isolated the fish population within the sampling site. The area of this closed waterway was measured using a laser distance meter. The captured fish biomass was recorded by repeated sampling (fourfold) of this closed population over short sequential time intervals (4 h). In brief, approximately 0.5 g of tail fin was collected from each captured specimen and preserved in a 95 % alcohol solution, after which the captured fish were released into the outer region of the enclosed waterway. Voucher specimens were deposited in the Natural Museum of the Northwest Institute of Plateau Biology, Chinese Academy of Sciences.

2.5. Estimating the local absolute biomass through CPUE data and the Leslie regression model

The Leslie regression model (Leslie and Davis, 1939) is commonly employed to estimate the stock biomass of aquatic organisms through removal. The assumptions of the model are as follows (Yüksel et al., 2013): 1) the population must be closed, and 2) the CPUE data are collected through repeated samplings in this closed population over

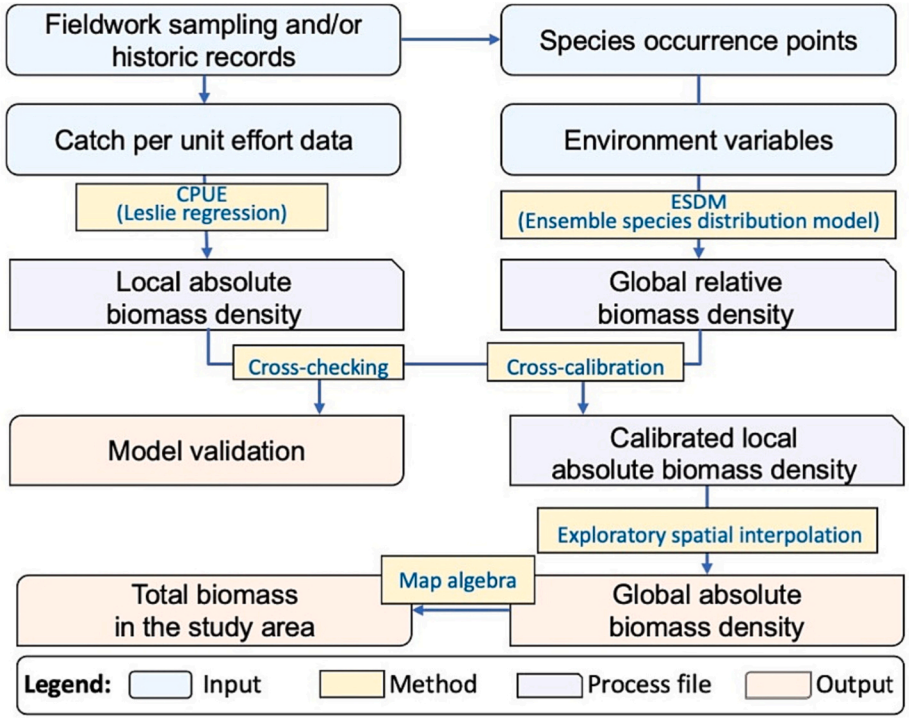


Fig. 1. Modeling framework of CPUESDM. CPUE: Catch per unit effort; ESDM: Ensemble species distribution modeling.

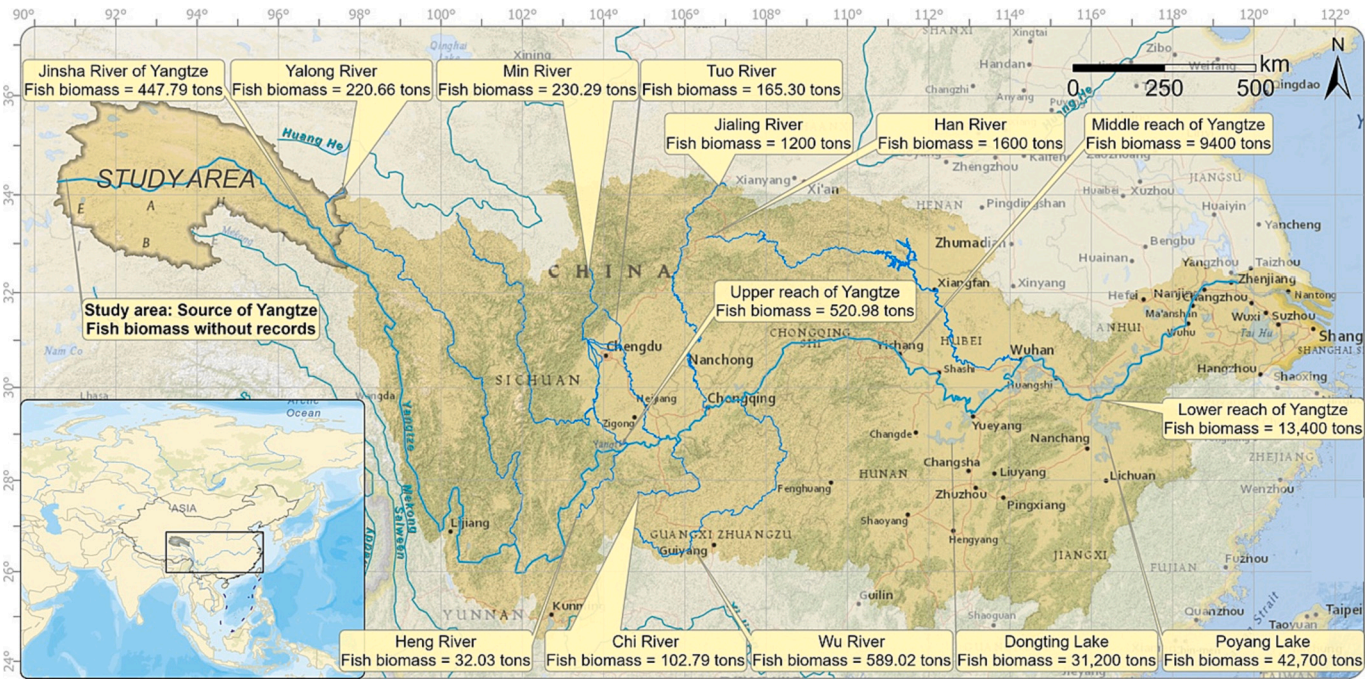


Fig. 2. Topographic map displaying the location of the study area in the Yangtze Basin. The fish biomass in 2020 in the Yangtze and its tributaries and lakes (data from Dong et al., 2023) is marked by boxes in yellow.

consecutive short time intervals, ensuring that the catchability of the targeted fish remains constant (i.e., equal amounts of effort are expended) during each sampling period. Additionally, fertility and mortality are negligible during these short sampling periods. The fieldwork sampling method outlined in Section 2.4 satisfies these assumptions.

The Leslie regression model is a removal method that involves sequential sampling of the target population, similar to the mark-recapture and maximum-likelihood methods. In each sampling period, the number of captured fish is recorded and temporarily removed from the population, resulting in a decrease in catch (i.e., CPUE) during subsequent sampling periods. The decline rate of CPUE provides a quantifiable measure of the proportion of the initial population biomass (Hayes et al., 2007). Although the Leslie regression model has more stringent assumptions than approaches like the maximum likelihood model or the mark-recapture method (Cowx, 1983), CPUESDM focuses on fish stock biomass assessment with limited data, where rigorous assumptions can minimize negative impacts from sampling. Therefore, we selected the Leslie regression model as the fundamental algorithm for CPUESDM.

The mathematic interpretations are as follows:

$CPUE_t$ is defined as the ratio of the catch (C_t) to effort (f_t) over time $t-1$ to t (Eq. (1)). Under the assumptions listed above, the catchability (q) is constant because the efforts of each catch during time interval t are identical. Therefore, $CPUE_t$ also represents the instantaneous catch biomass at time t, which is proportional to population size (N_t) and catchability (q) (Eq. (1)).

$$CPUE_t = \frac{C_t}{f_t} = qN_t \tag{1}$$

Subsequently, the original population size without catching is represented by N_0 (i.e., $t = 0$), and K_t is defined as the cumulative catch from time 0 to the start of time t. Thus, the population size at time t (N_t) is the difference between N_0 and K_t (Eq. (2)).

$$N_t = N_0 - K_t \tag{2}$$

Therefore, substituting Eq. (2) for N_t in Eq. (2) gives you Eq. (3), or the Leslie Regression Model,

$$CPUE_t = \frac{C_t}{f_t} = qN_0 - qK_t \tag{3}$$

where catchability (q) is constant, $CPUE_t$ and N_0 are linearly correlated. This linear regression equation can be determined utilizing the least-squares method, where the slope $= -q$, and the Y-axis intercept $= qN_0$. The local absolute biomass (original population size, N_0) can be estimated from this regression line using Eq. (4).

$$N_0 = \frac{qN_0}{q} = -\frac{\text{Y axis intercept}}{slope} = \text{X axis intercept} \tag{4}$$

As t approaches a sufficiently large value in Eq. (3), such that $CPUEt$ tends to zero, the X-axis intercept represents $N_0 = K_t$. The parameters of the Y-axis intercept (qN_0) and the slope ($-q$) can be calculated using the least-squares method. Alternatively, CPUESDM enables the splitting of input CPUE data into training and testing datasets. Subsequently, a cross-validation analysis was conducted between the training and testing datasets to evaluate the model accuracy. This alternative approach is recommended when $t > 20$.

After determining N_0, the local absolute population density of species j at sampling site i ($Dlocal_{ij}$) was calculated as the quotient of N_{0ij} and the area of the closed population s_i (Eq. (5)). If the $Dlocal_{ij}$ values at all sampling sites conformed to a normal distribution by Shapiro-Wilk test, the influence of seasonal variations in fish distribution on the outcomes could be disregarded.

$$Dlocal_{ij} = \frac{N_{0ij}}{s_i} \tag{5}$$

The case study retained 15 groups (i.e., i ranging from 1 to 15) of the CPUE data with a coefficient of determination $r^2 > 0.8$ for further analysis (Guo et al., 2022). This modeling process was carried out using Python codes scripted using the Scikit-learn package v1.3.0 (a machine learning package in Python) (Abraham et al., 2014) and the geometry token functions of the ArcPy site package (v3.1) (Zandbergen, 2020). Section 3.1 of Supplementary material 2 describes the modeling processes of this step.

2.6. Estimating global relative biomass by ESDM

The ESDM method estimates the global relative biomass ($Hglobal_{ij}$) using limited species occurrence data and the corresponding abiotic/environmental data as input. The $Hglobal_{ij}$ indicator ranges from zero to one, with zero representing the minimum suitable habitat and one indicating the most suitable habitat. In this case study, the ESDM comprised 106 occurrence points (Table S2 in Supplementary material 1) obtained from official fishery department records and our long-term fieldwork monitoring conducted between 2005 and 2022. The occurrence data were thinned using the spThin package with a thinning parameter of 10 km to mitigate potential spatial autocorrelation. After thinning, 105 occurrence points remained for further analysis. Subsequently, climate and elevation data [Shuttle Radar Topography Mission (SRTM)] were downloaded from WorldClim (WorldClim.org). Flow accumulation, direction, and Strahler stream orders were derived from the SRTM elevation layer using ArcGIS Pro. All layers were clipped according to the boundaries of the study area (Fig. 2). Next, a multicollinearity test was performed on all variables, and those with a Pearson's coefficient > 0.75 were removed to avoid collinearity of variables (Fig. S2 in Supplementary material 1). Only flow accumulation, flow direction, Strahler stream order, bio2, bio11, bio13, and bio14 were used in the modeling process. Ten robust or machine learning algorithms were utilized, including the generalized linear model, generalized boosted model, generalized additive model, classification tree analysis, multivariate adaptive regression spline, artificial neural network, surface range envelope, flexible discriminant analysis, random forest, and maximum entropy algorithms. One thousand randomly generated pseudo-absence locations were used. Furthermore, we split the dataset into two 3:7 ratios for model calibration and validation. Receiver operating characteristic (ROC) curve analysis was used to evaluate the models, and a threshold ROC >0.7 was required for models to be included in the ensemble model (Ahmad et al., 2020). This modeling process was implemented using the Biomod2 R package (Thuiller et al., 2013), and the source codes are presented in Section 3.2 of Supplementary material 2.

2.7. Cross-checking and cross-calibration

A generalized linear model was employed to cross-check and cross-calibrate the independent calculations. *Hglobal* is a simulated value, while *Dlocal* represents the observed value. After obtaining Cohen's kappa coefficient between the max-min normalized simulated and observed values, the consistency between the simulated and observed values were used to validate model accuracy according to the following scale: 0–0.20, extremely weak consistency; 0.21–0.40, weak consistency; 0.41–0.60, moderate consistency; 0.61–0.80, good consistency; 0.81–1.00, extremely good consistency (Rigby, 2000). Additionally, the predicted values of the generalized linear model from the cross-calibration between $Dlocal_{ij}$ and $Hglobal_{ij}$ represent the calibrated local absolute density ($Dlocal\_calibration_{ij}$), thereby mitigating the impact of limited data on the results.

The source codes, as viewed in Section 3.3 of Supplementary material 2. The generalized linear regression was carried out using the "stats" class of the ArcPy site package (Zandbergen, 2020).

2.8. Exploratory spatial interpolation

The spatial interpolation method is commonly used to identify spatiotemporal patterns of fishery resources (Alam et al., 2016; Yin et al., 2022), which produces a range of interpolation outcomes based on the input point features and a field (i.e., $Dlocal\_calibration_{ij}$), which are subsequently evaluated and ranked using a customizable criterion grounded in cross-validation statistics. CPUESDM employs 20 customized spatial interpolation methods to identify the method with the highest prediction accuracy. The list of these 20 spatial interpolation methods and their descriptions and parameters are listed in Table S3 of Supplementary material 1. This exploratory spatial interpolation process was conducted using a Python script, as interpreted in Section 3.3 of Supplementary material 2. The exploratory interpolation and geostatistical layer to raster and contour functions in the Geostatistical Analyst class of the ArcPy site package (Zandbergen, 2020) were employed to implement this analysis.

2.9. Extracting global absolute biomass

Step 2.8 translates the calibrated local absolute density ($Dlocal\_calibration_{ij}$) to the global absolute density ($Dglobal_{ij}$). As aquatic areas spatially restrict the distribution of fish biomass, the output of step 2.8 (i.e., $Dglobal_{ij}$) was extracted from the focused aquatic areas in the study region (Eq. (6)), resulting in the biomass density in the aquatic areas of species j at cell i ($Dglobal\_water_{ij}$). Finally, the total biomass of species j was equal to the sum of biomass density $Dglobal\_water_{ij}$ multiplied by the cell area (Eq. (7)). A map-algebra algorithm was used to complete the calculation. The Python source codes are explained in Section 3.3 of Supplementary material 2.

$$Dglobal\_water_{ij} = Dglobal_{ij} \times \frac{S_{water_i}}{CellSize^2} \tag{6}$$

$$Bglobal\_water_{jtotal} = \sum\nolimits_{i=0} Dglobal\_water_{ij} \times CellSize^2 \tag{7}$$

In Eq. (6) $Dglobal\_water_{ij}$ represents the biomass density in the aquatic areas of species j in cell i. S_{wateri} represents the aquatic area of cell i. In Eq. (7), $Bglobal\_water_{jtotal}$ is the total biomass in the study area of species j.

The parameter cell size is related to the simulation accuracy; here the cell size in Eq. (6) is the raster size of the $Dglobal_{ij,}$ which is determined by the modeling environment. The cell size in Eq. (7) is a contrast determined by user input, i.e., the contrast cell size used here was 100 m.

2.10. Sensitivity analysis

Reducing the parameter cell size leads to increased accuracy but also exponentially increases the processing time of the computer. Therefore, a sensitivity analysis was conducted to test the sensitivity of $Bglobal\_water_{jtotal}$ to increasing cell size. By performing 991-fold repetitions of Step 2.9 while varying the parameter cell size to an integer from 10 to 1000 with a for-loop structure, a single-sample t-test was performed to compare different cell size ranges and to determine an optimal value that balances the processing time and acceptable accuracy.

3. Results

3.1. Local absolute biomass density (Dlocal)

CPUESDM script 01 (Section 3.1 of Supplementary material 2) employs the Leslie regression model (Section 2.3) to estimate local absolute biomass density (*Dlocal* in (Eq. (5)). The input consists of a point .shp file representing the sample site locations, original CPUE data (*CPUEt* in Eqs. (1) and (3)), and closed population area (s in m^2 in (Eq. (5)). The output (Table S4 of Supplementary material 1) includes the cumulative catch from time 0 to the beginning of t (*Kt* in Eq. (2)), the initial population size without any fishing activity in this closed population (N_0 in Eq. (4)), the coefficient of determination (r^2) between *CPUEt* and *Kt* through least-squares analysis, the mean standardized error, and *Dlocal*.

In the case study, 15 groups of CPUE data with a coefficient of determination (r^2) >0.8 were selected for further analysis (Fig. S3 of Supplementary material 1). The *Dlocal* ranged from 0.051 to 0.221 g/m^2, with a mean value of 0.173 g/m^2 and a standard deviation (STD) of 0.052 ($n = 15$). According to the Shapiro-Wilk test, the 15 *Dlocal* data points did not conform to a normal distribution ($W = 0.789$, $p = 0.003$),

which suggests that the effect of seasonal variability on fish distribution may have led to an overestimation. As all CPUE data were collected during the summer, the target species likely migrated exclusively from downstream areas to the sampling sites for spawning and breeding purposes during this period. Consequently, the findings solely reflect biomass levels during the breeding season within the study area. If $Dlocal_{ij}$ values at all sampling sites followed a normal distribution pattern, the influence of seasonal variations on fish distribution would have been disregarded.

3.2. Global relative biomass density (Hglobal)

The CPUESDM script 02 (Section 3.2 of Supplementary material 2) utilizes 10 robust or machine-learning ESDM algorithms (Section 2.6) derived from the Biomod2 R package (Thuiller et al., 2013) to estimate the global relative biomass density. The input consists of species occurrence points (Table S2 of Supplementary material 1) and corresponding environmental variables (Fig. S2 of Supplementary material 1). The output is a raster file representing the global habitat suitability within the study area, *Hglobal* values, as detailed in Section 2.6.

3.2.1. Model performance

The ESDM results demonstrated satisfactory internal evaluation on average (Table 1), with the sensitivity and specificity meeting the performance requirements of the model. The calibration accuracy was determined to be 88.0 % via cross-validation with a 3:7 test-to-training data ratio. These findings indicate that the model exhibits acceptable accuracy and reliability.

3.2.2. Global relative biomass density and its spatial distribution characteristics

The global relative biomass density map (Fig. 3b) comprised 8902 cells within the aquatic area (mean = 0.31, STD = 0.300, cell size = 1 km^2). The middle reaches of the study area exhibited relatively high habitat suitability. The results obtained from the hot spot analysis (Getis-Ord Gi analysis) (Getis and Ord, 1992), as shown in Fig. 4, indicated that areas with high *Hglobal* values were concentrated midstream of the study area, including in certain adjacent tributaries, such as the Dam and Lema rivers (hot point with confidence coefficient = 99 % in Fig. 4). Furthermore, the *Hglobal* values displayed a negative correlation with both human footprint and altitude and a positive correlation with the profile curvature of the river (Fig. 5), suggesting that *H. microcephalus* prefers habitats characterized by rapidly flowing water at lower altitudes while actively avoiding areas with high levels of human activity.

3.3. Model validation

The initial stage of CPUESDM script 03 (Section 3.3 of Supplementary material 2) involved the application of a generalized linear model to perform cross-checking and cross-calibration between *Hglobal* and *Dlocal* (points in Fig. 4 and Table 2), as described in Section 2.7. The results from the cross-check indicated a Cohen's kappa coefficient of 0.78, demonstrating strong agreement between the max-min normalized simulated and observed values. Additionally, the cross-calibration analysis revealed residuals ranging from −0.004 to 0.006 g/m^2 between the *Dlocal* and *Dlocal_calibration* values. By employing this generalized linear model, we derived a calibrated local biomass density, denoted as *Dlocal_calibration*, which was further utilized to improve the accuracy of the results when working with limited data.

Table 1
Model performance of ten ensemble species distribution modeling for the case study.

Metric	Algorithm	Cutoff	Sensitivity	Specificity	Calibration
ROC	Mean	363.50	95.65	76.61	0.88

3.4. Global absolute biomass density (Dglobal)

The subsequent phase of CPUESDM script 03 (Section 3.3 of Supplementary material 2) implements an exploratory spatial interpolation function to convert the localized point values (i.e., *Dlocal_calibration*) into global polygon values (i.e., *Dglobal*), as discussed in Section 2.8. Finally, CPUESDM script 04 (Section 3.4 of Supplementary material 2) extracts the *Dglobal* values for the designated aquatic region following the guidelines outlined in Section 2.9.

The results of the case study (Table S5 of Supplementary material 1) demonstrate that, of the 20 exploratory spatial interpolation methods, the simple kriging method with default parameters exhibited superior performance, verified by having the lowest root mean square error (RMSE). A smaller RMSE value indicates higher accuracy, typically below 0.50, for a robust modeling performance. Our case study revealed exceptional model accuracy using a simple kriging method with default parameters, yielding an RMSE of 0.05. Consequently, we selected this result (Fig. 6a) for further analysis.

The exploratory spatial interpolation generated the *Dglobal* value raster map, as illustrated in Fig. 6a. Overall, the distribution of *Dglobal* areas exhibited a concentration in the middle reaches of the study area (Fig. 6a) and at altitudes ranging from 4300 to 5100 m (Fig. 6b). The total fish biomass ($Bglobal\_water_{jtotal}$ in Eq. (7)) in the Tuotuo, Yangtze, Chumar, and Dam rivers was calculated to be 113.52 tons (mean = 0.192 g/m^2) (Table 3). Locally, a cold center and a hot center were identified (Fig. 6a).

3.4.1. Zonal statistics by Dglobal cold and hot centers

The *Dglobal* cold center (Fig. 6a) was located at point 10 in Fig. 3b and its surrounding areas, where both *Dlocal* (0.051 g/m^2) and *Hglobal* values (0.23) exhibited relatively minor magnitudes across the study area (Table 2). Fig. 6c illustrates the representative braided river fish habitat within the cold center, characterized by a network of multiple shallow channels that bifurcate and converge. In this particular region, the stream gradient (average profile curvature = 0 %, n = 58,054, STD = 0.000263) exhibits a comparatively milder nature than in the upstream gorge areas, leading to a reduction in water velocity and consequent augmented deposition of sediment. The presence of a bridge spanning the Chumar River (Fig. 6d), accompanied by adjacent petroleum pipelines, partially isolates fish habitats. Consequently, considering these factors collectively, the cold center was identified as a priority area for implementing diverse measures to restore habitats, such as eco-shoreline development, fish passage facilities, and artificial fish nests.

Comparatively, the *Dglobal* hot center (Fig. 6a) was identified at points 2, 3, 5, 6, 8, and 9 in Fig. 3b, along with their respective buffer zones. The weighted average value of *Dlocal* (weighted by *s*) was determined to be equal to 0.214 g/m^2, and the *Hglobal* values were recorded as an average of 0.95 (Table 2). Fig. 6e depicts a representative river valley fish habitat situated within the hot center, characterized by a higher stream gradient (average profile curvature = −0.061 %; n = 1,680,025; STD =0.000783), deeper channels without bifurcation, and more rapid water velocity than those of the cold center. Although these aquatic areas cover only 12.76 % of the total water surface area (403.29 km^2) among the Yangtze River sources, they play a crucial role in supporting a significant fish biomass of 14.85 tons accounting for 13.08 % of the total biomass (113.52 tons) (Table 3). Consequently, it is imperative to recognize hot centers as core habitats for conservation purposes and accordingly prioritize fish stocking efforts. This approach was designed to minimize the adverse effects of human activities on riparian zones and water quality while mitigating potential disruptions caused by dams or other structures that impede the longitudinal connectivity of rivers.

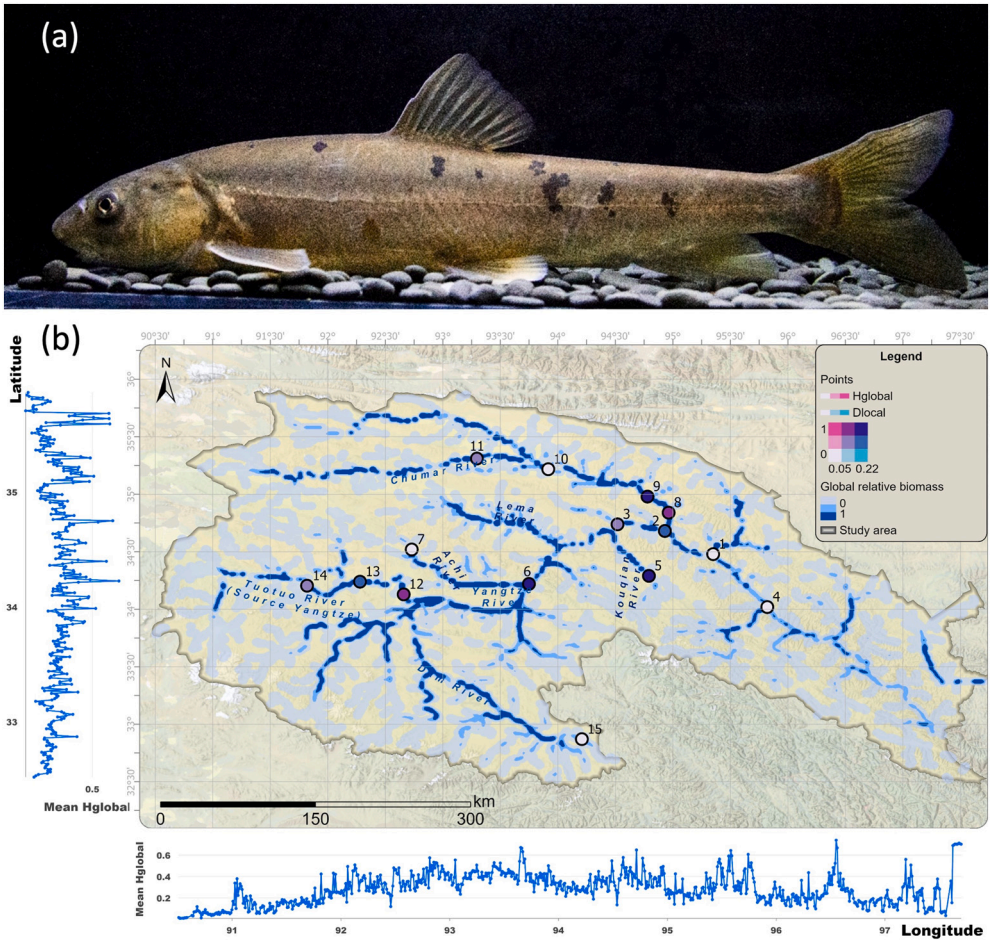


Fig. 3. (a) Image of the targeted species *Herzensteinia microcephalus*, sampled at point 13 in panel b, captured by the authors. (b) Global relative biomass density (*Hglobal*) and local absolute biomass density (*Dlocal*). The points represent the CPUE sampling sites, and their ID corresponds to those in Fig. S3, Table S4, and Table 2.

3.5. Sensitivity analysis

The parameter cell size for the map algebra calculation (Section 2.9) is determined subjectively using user input. Smaller cell sizes are associated with higher accuracy; however, the computational processing time increases exponentially as cell size decreases. Therefore, it is essential to explore an appropriate cell size for model testing through a sensitivity analysis. Fig. 7 and Table 4 illustrate the sensitivity analysis of 991 iterations between cell size (m) and total biomass ($Bglobal\_water_{total}$, tons). The results indicated that when 10 m $\leq$ cell size $\leq$100 m, an STD of 0.120 was observed with a significant *p*-value; however, for cell sizes ranging from 100 to 1000 m, the STD increased to 2.123, with a *p*-value that was statistically insignificant. These results suggest that the selected cell size of the case study of 100 m is justifiable. The computer processing time for this case study using a Mac Pro 2019 graphic workstation (Apple Inc., Cupertino, California, USA) with standard configuration was measured to be approximately 243 s when employing a cell size of 100 m. However, when utilizing a smaller cell size of 10 m, there was a significant increase in the processing time to approximately 1.53 h. Consequently, it could be inferred that opting for a cell size of l00 m aids in achieving an optimal balance between processing efficiency and accuracy in our analysis.

4. Discussion

4.1. Comparison between different methods

The methods commonly used for the stock biomass estimation of aquatic organisms can be categorized into four groups (Table 5): capture-based methods (Hayes et al., 2007), eDNA analysis (Doi et al., 2015), acoustic monitoring (Block et al., 2019), and removal techniques (Hayes et al., 2007). Table 5 lists illustrative examples and comprehensive descriptions of these methods. Although capture-based methods are extensively utilized, they rely heavily on high-frequency fish catch data, rendering them more suitable for commercial fisheries than for conservation efforts targeting endangered wild fish populations. Conversely, while eDNA analysis offers a non-intrusive approach, the uncertainty surrounding the correlation between eDNA concentration and organism biomass poses challenges in translating relative biomass into absolute values (Li et al., 2020). Furthermore, the implementation of acoustic monitoring using echo-sounders or sonar requires navigable water or fixed echo-sounding platforms, potentially limiting its practicality in certain highland freshwater regions.

Therefore, CPUESDM utilizes the Leslie regression model (Leslie and Davis, 1939) as a crucial step to capture relative biomass using CPUE data and subsequently convert it into absolute biomass using the least-squares method. This case study also demonstrated a strong agreement between the results obtained from Leslie's method and the ESDM (Cohen's kappa coefficient = 0.78, Section 3.3). Furthermore, compared to Leslie regression methods, the maximum-likelihood method offers

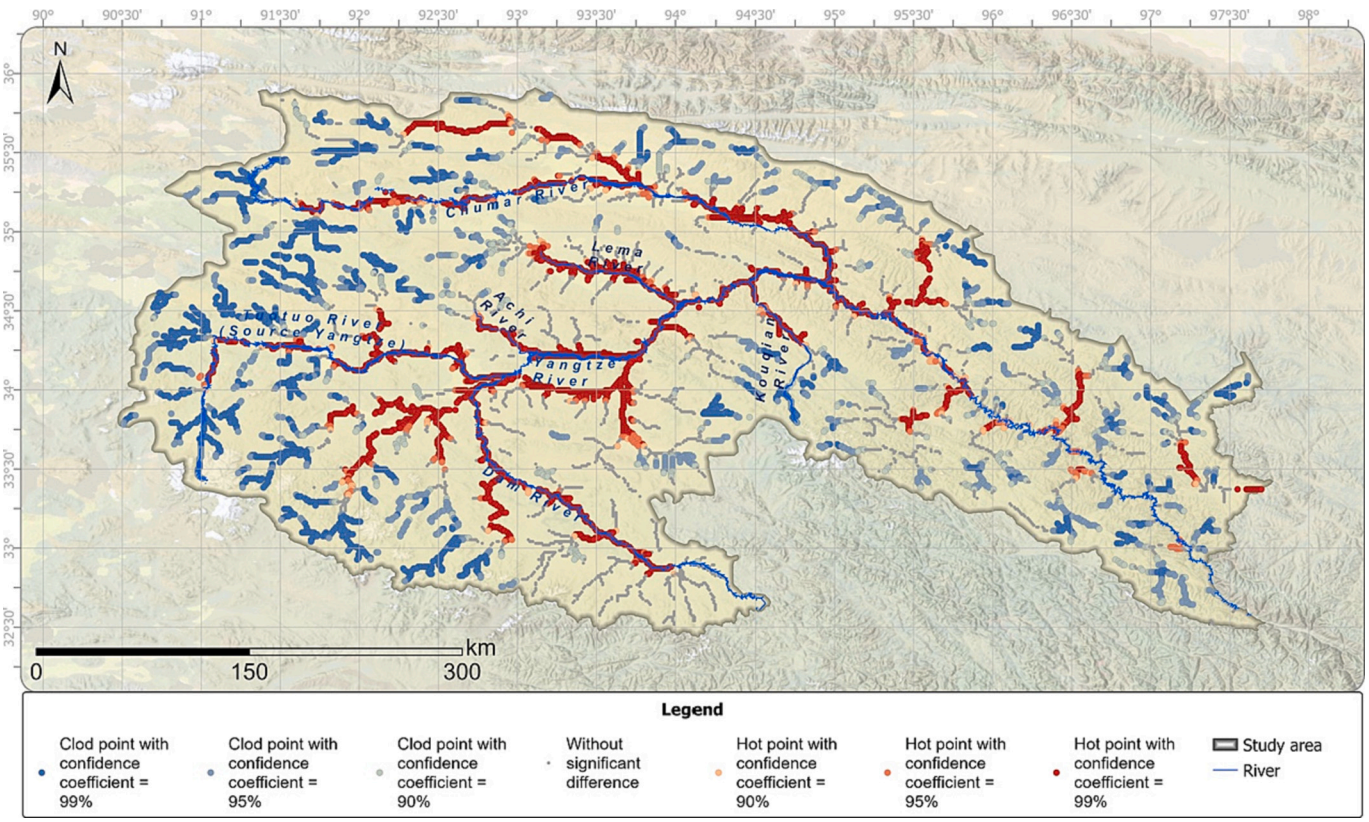


Fig. 4. Hotspots (red) and cold spots (blue) of global relative biomass density.

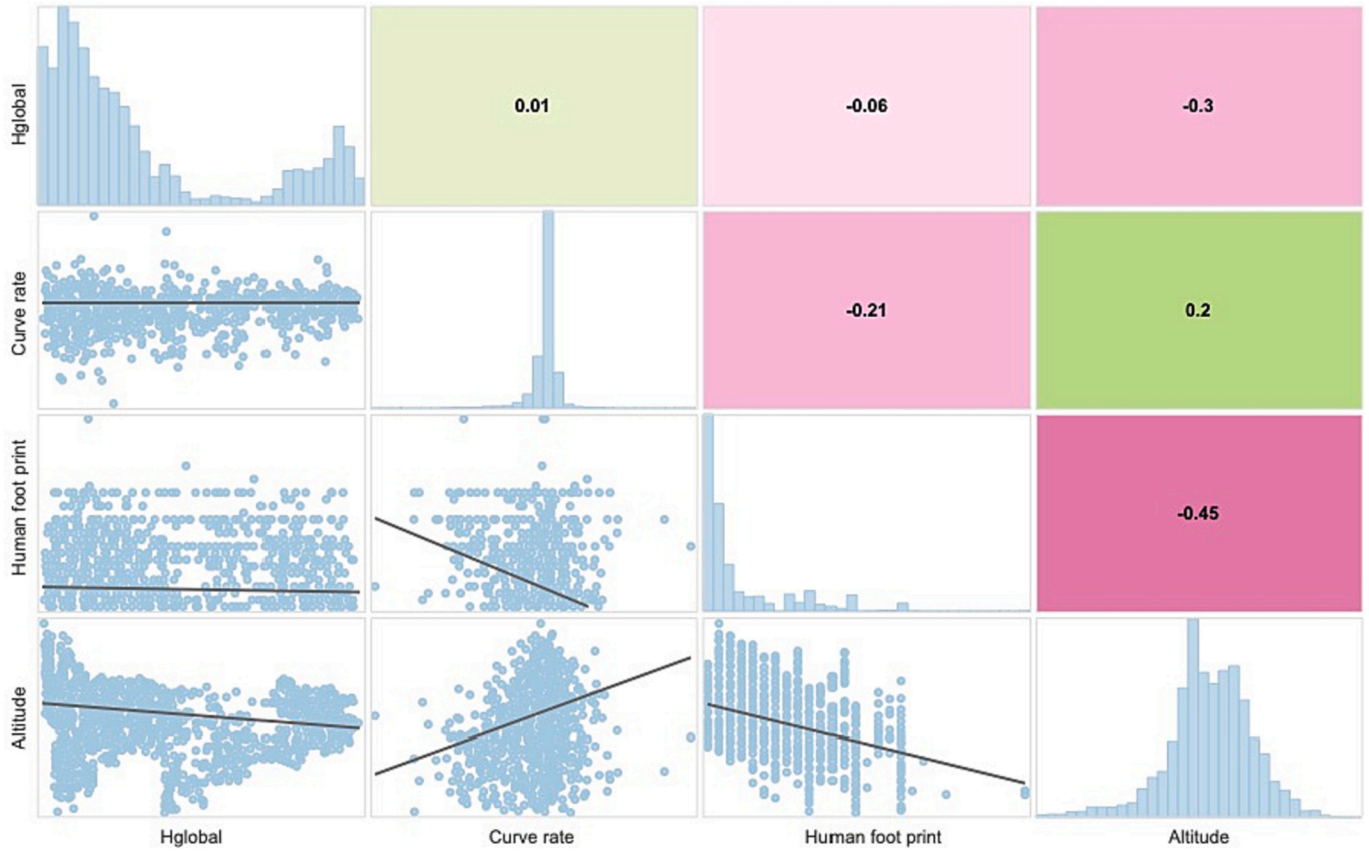


Fig. 5. Pearson correlation matrix among the global relative biomass density (*Hglobal*), profile curvature of the river (curve rate), human footprint, and altitude. Green refers to a positive correlation, while red represents a negative correlation.

Table 2
Local absolute biomass density (*Dlocal*) and its calibration value by the generalized linear model (*Dlocal_calibration*), global relative biomass density (*Hglobal*), the residual between *Dlocal* and *Dlocal_calibration*, and the standardized residuals (STD Residual).

ID	*Dlocal* (g/m^2)	*Hglobal*	*Dlocal_calibration* (g/m^2)	Residual (g/m^2)	STD residual
1	0.170	0.76	0.171	0.000	−0.098
2	0.208	0.91	0.204	0.004	1.433
3	0.189	0.86	0.193	−0.004	−1.190
4	0.167	0.76	0.170	−0.003	−1.124
5	0.208	0.94	0.210	−0.001	−0.382
6	0.221	0.98	0.220	0.002	0.585
7	0.092	0.41	0.091	0.000	0.159
8	0.204	0.93	0.207	−0.003	−0.963
9	0.218	0.97	0.218	0.001	0.183
10	0.051	0.23	0.051	0.000	−0.047
11	0.188	0.83	0.186	0.002	0.550
12	0.203	0.93	0.208	−0.004	−1.432
13	0.207	0.90	0.200	0.006	2.090
14	0.177	0.79	0.177	0.000	−0.008
15	0.092	0.41	0.091	0.001	0.244

significant advantages, as it enables testing of assumptions related to removal methods and the creation of models that accommodate a more relaxed set of assumptions (Hayes et al., 2007). However, to successfully implement the maximum-likelihood method, a minimum of 20 instances of sampling over short sequential time intervals ($t > 20$ in Eqs. (1), (2), and (3)) are required, as suggested by Abraham et al. (2014). However, our fieldwork (Section 2.4 and Table S4) revealed that four instances of short sequential time interval sampling were adequate to fulfill the rigorous assumptions required by the Leslie regression model (Section 2.5). Compared to the maximum-likelihood method, which can be negatively affected by limited data, the Leslie regression model is more suitable for minimizing such effects when working with endangered species.

4.2. Comparison of waters adjacent to the study area

Our results indicate that the stock biomass of *H. microcephalus* in the source of Yangtze was estimated to be 113.52 tons (Table 3). Based on our unpublished extensive fieldwork data collected from 2005 to 2022, *H. microcephalus* accounted for approximately 37.81 % of the relative importance index of all species in the study area. Consequently, we estimated a total fish stock biomass of approximately 300.24 tons by dividing 113.52 tons by 37.81 %. This finding aligned with adjacent waters within our study area, where it reported that total fish stock biomasses were recorded as 400.74 tons for Jinsha River and 520.98 tons for the upper reaches of Yangtze (Fig. 2) (data from Dong et al., 2023). Therefore, this comparison validates the reasonableness and comparability of the case study results.

4.3. Conservation implications of H. microcephalus

Notably, the fish stock biomass data presented in Fig. 2 were assessed using a CPUE method based on extensive historical catch records before implementing the "Ten-Year Fishing Moratorium in the Yangtze Basin" conservation project. This project prohibited commercial fishing activities in key waters along the Yangtze River, resulting in a lack of available catch records. Therefore, our findings serve as baseline data to assess the effectiveness of protection measures following the completion of this conservation project. Meanwhile, we suggest a wider adoption of the CPUESDM method to compensate for the limited fish catch data during this ongoing conservation project. For example, identifying *H. microcephalus* biomass hotspots provides valuable information for prioritizing conservation efforts and targeted species stocking. Furthermore, the identification of biomass cold spots (Fig. 6a) indicates potential habitat fragmentation resulting from human activities, necessitating appropriate habitat restoration measures.

4.4. Innovations, limitations, and future study

The innovations of CPUESDM include:

1) By utilizing only CPUE data (Table S4 of Supplementary material 1) and a limited number of occurrence points (Table S2 of Supplementary material 1), the adverse effects of extensive sampling on endangered fish populations can be minimized. Additionally, this approach is particularly suitable for mountainous freshwater habitats lacking navigable conditions. Even in cases where CPUE data are not readily available, alternative methods, such as experimental eDNA data or acoustic monitoring data, can serve as viable substitutes.
2) The modeling processes of the CPUE (Section 2.5) and ESDM (Section 2.6) are conducted independently, enabling cross-checking to verify the accuracy of the CPUESDM simulation results (Section 2.7).
3) The final output is a raster map that explicitly depicts the spatial distribution of the fish stock biomass in each cell (Fig. 6a). In contrast with other models that solely provide global aggregate outcomes, this model more accurately portrays the spatial heterogeneity of fish stock biomass, as exemplified by the presence of hot and cold centers in Fig. 6a.

After hundreds of modeling tests, the currently known limitations to the method used here are as follows: First, compared with that of other methods (listed in Table 5) that utilize additional input sampling data, CPUESDM may exhibit relatively lower accuracy. However, CPUESDM focuses on limited data concerning freshwater fish species. Second, the assumptions required by the Leslie regression model are stringent (Section 2.5) but can be met using specific sampling methods, such as those used in this study. Finally, relying solely on CPUE data might overlook the influence of weather and seasonal variability on fish distribution. For example, our case study exclusively reflects biomass levels during the breeding season but potentially overestimates the average annual level.

In future studies, we will employ the CPUESDM model to elucidate the potential acceleration of niche divergence among sympatric fish species on the Tibetan Plateau (Jia et al., 2020), a phenomenon that is potentially leading to a substantial reduction in habitat availability for *H. microcephalus*. Furthermore, we will continuously refine and enhance the methodology based on valuable user feedback.

5. Conclusions

In this study, we introduced CPUESDM, an innovative model framework designed to enable spatially explicit estimation and validation techniques for assessing the stock biomass of freshwater fish species using limited data and applied it to a case study on *H. microcephalus*, an endangered endemic fish in the Yangtze River Basin. A detailed step-by-step instruction manual is provided in the Supplementary material to facilitate users in replicating CPUESDM using their self-fieldwork data. Overall, the findings of novel combined approach offer pivotal insights for effective conservation management, identifying priority areas for conservation, and selecting suitable locations for targeted fish stocking across fish systems.

Funding

This research was funded by Natural Science Foundation of Qinghai Province (2022-ZJ-936Q), China Postdoctoral Science Foundation No. 2021M693373, the Projects Investigation of Aquatic Biological Resources supported by Department of Agriculture and Rural Affairs of Qinghai Province No. E039831D01 & No. E239421D01, and Qinghai

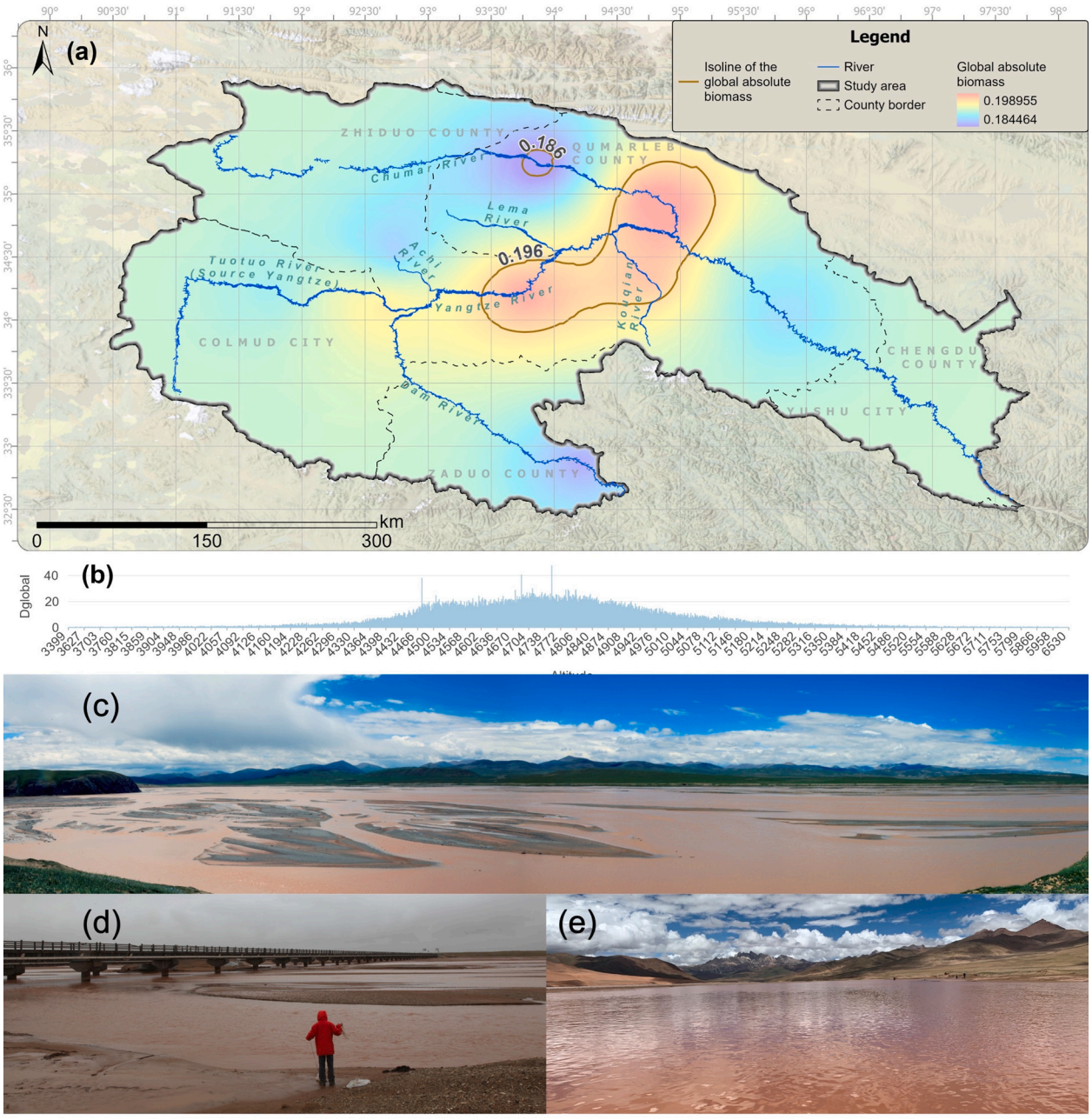


Fig. 6. (a) Global absolute biomass density (*Dglobal*, g/m^2) raster map. The red circle is the hot center with a *Dglobal* value >0.196 g/m^2, whereas the blue circle is the cold center with a *Dglobal* value <0.186 g/m^2. (b) Relationship between the accumulated *Dglobal* and altitude. (c) Representative braided river fish habitat in the cold center. (d) Fish habitat affected by human construction in the cold center, where the Chumar River is isolated by a bridge of the G215 national highway and the adjacent petroleum pipelines. (e) Typical river valley fish habitat in the hot center.

Table 3
Aquatic area (km^2), sum biomass of *H. microcephalus* (*Bglobal_water*$_{jtotal}$, ton), and global absolute biomass density (*Dglobal_water*, g/m^2) in the designated aquatic region[a] of the hot and cold centers.

Type	Aquatic area	Area%	*Bglobal_water*$_{jtotal}$	*Bglobal_water*$_{jtotal}$%	*Dglobal_water*
Cold	14.82	3.68 %	4.11	3.62 %	0.185
Hot	51.46	12.76 %	14.85	13.08 %	0.197
Others	337.01	83.56 %	94.56	83.30 %	0.191
Sum	403.29[a]	100.00 %	113.52	100.00 %	0.192[a]

[a] Note: The designated water region for statistical analysis encompasses the Tuotuo, Yangtze, Chumar, and Dam rivers (Fig. 6a), with a combined area of 403.29 km^2. The average global absolute biomass density weighted by aquatic area was 0.192 g/m^2.

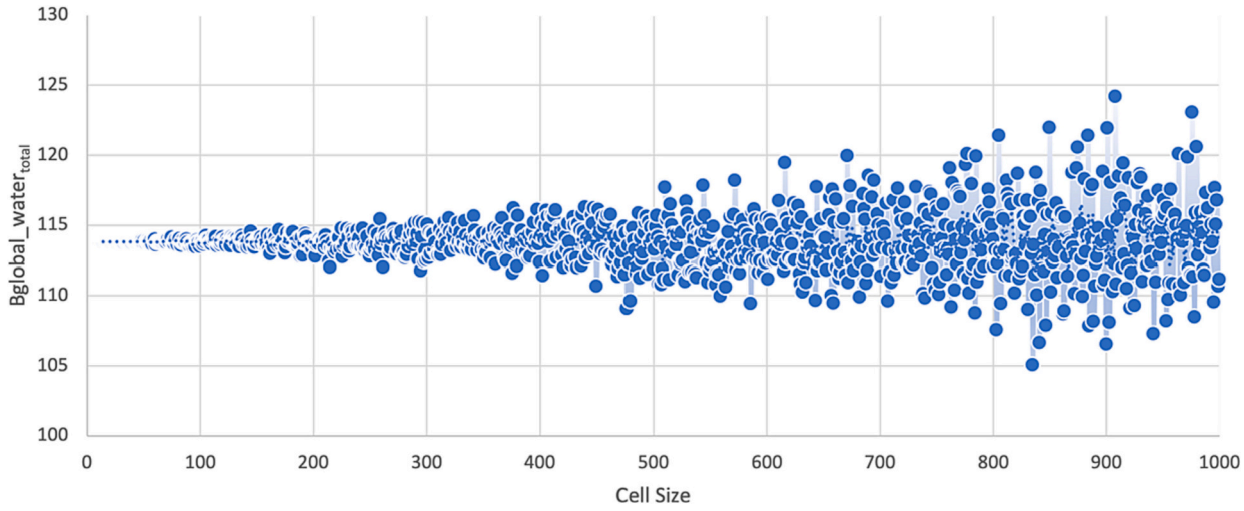


Fig. 7. Sensitivity between cell size (m) and the total biomass of *H. microcephalus* (*Bglobal_water*$_{total}$, tons) in the Tuotuo, Yangtze, Chumar, and Dam rivers.

Table 4
Results of the single sample *t*-tests between 10 ≤ cell size ≤100 and 100<cell size ≤1000 compared to the mean output *Bglobal_water*$_{total}$ = 113.84 of the 991 iterations.

Samples	*n*	Min	Max	Average	STD	*p*
10 ≤ cell size ≤ 100	91	113.42	114.15	113.81	0.120	0.024*
100<cell size ≤ 1000	900	105.09	124.20	113.85	2.123	0.924**

* $p < 0.05$.
** $p < 0.01$.

Table 5
Certain commonly used methods for biomass estimation of aquatic organisms.

ID	Method categories	Examples	Descriptions
1	Capture-based method (Hayes et al., 2007)	1) Count directly 2) Length-based fish stock assessment 3) Mark–recapture method	High accuracy, but the amount of fish catching is large. Usually suitable for commercial fishery production.
2	eDNA (Doi et al., 2015)	1) Quantitative real-time polymerase chain reaction 2) Droplet digital polymerase chain reaction	Non-intrusive to targeted fish species, but difficult to translate the relative biomass to absolute biomass (Li et al., 2020).
3	Acoustics monitoring (Block et al., 2019)	1) Navigational echo-sounding 2) Fixed echo-sounding	Non-intrusive to targeted fish species, but usually feasible in navigable waters for large-scale monitoring or require a fixed echo-sounding platform for small-scale monitoring.
4	Removal method (Hayes et al., 2007)	1) Least-squares method (Leslie regression model) 2) Maximum-likelihood method	Sampling at short sequential time intervals. The absolute biomass can then be measured using the relative biomass.

Kunlun Ying Cai Action Project No. Qing Ren Zi [2020] 18.

CRediT authorship contribution statement

Kunyuan Wanghe: Conceptualization, Writing – original draft. **Shahid Ahmad:** Methodology. **Xin Zhou:** Visualization. **Fei Tian:** Funding acquisition. **Sijia Liu:** Project administration. **Bingzheng Zhou:** Investigation. **Ghulam Nabi:** Supervision. **Guojie Wang:** Resources. **Kemao Li:** Supervision. **Shenglong Jian:** Data curation. **Huamin Jiang:** Project administration. **Shengxue Chen:** Data curation. **Yimeng Niu:** Investigation. **Muhammad Ismail Khan:** Resources. **Kai Zhao:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.
Data for: Spatial explicit estimation of endangered freshwater fish biomass with limited data, as a case study of an endemic fish on the Tibet Plateau, China (Original data) (Mendeley Data)

Acknowledgements

We thank for the helps from the fishery administration departments for our long-time fish monitoring from 2005 to 2022.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2023.168717.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., Varoquaux, G., 2014. Machine learning for neuroimaging with scikit-learn. Front. Neuroinform. 8 (FEB) https://doi.org/10.3389/fninf.2014.00014.

Ahmad, S., Yang, L., Khan, T.U., Wanghe, K., Li, M., Luan, X., 2020. Using an ensemble modelling approach to predict the potential distribution of Himalayan gray goral (Naemorhedus goral bedfordi) in Pakistan. Glob. Ecol. Conserv. 21 https://doi.org/10.1016/j.gecco.2019.e00845.

Alam, R.Q., Benson, B.C., Visser, J.M., Gang, D.D., 2016. Response of estuarine phytoplankton to nutrient and spatio-temporal pattern of physico-chemical water quality parameters in Little Vermilion Bay, Louisiana. Ecol. Inform. 32 https://doi.org/10.1016/j.ecoinf.2016.01.003.

Arreguín-Sánchez, F., 1996. Catchability: a key parameter for fish stock assessment. Rev. Fish Biol. Fish. 6 (2) https://doi.org/10.1007/bf00182344.

Block, B.A., Whitlock, R., Schallert, R.J., Wilson, S., Stokesbury, M.J.W., Castleton, M., Boustany, A., 2019. Estimating natural mortality of Atlantic Bluefin tuna using acoustic telemetry. Sci. Rep. 9 (1) https://doi.org/10.1038/s41598-019-40065-z.

Chen, Z., Ren, Q., Liu, C., Xian, W., 2022. Seasonal and spatial variations in fish assemblage in the Yangtze Estuary and adjacent waters and their relationship with environmental factors. J. Mar. Sci. Eng. 10 (11) https://doi.org/10.3390/jmse10111679.

Cowx, I.G., 1983. Review of the methods for estimating fish population size from survey removal data. Aquac. Res. 14 (2) https://doi.org/10.1111/j.1365-2109.1983.tb00057.x.

Doi, H., Uchii, K., Takahara, T., Matsuhashi, S., Yamanaka, H., Minamoto, T., 2015. Use of droplet digital PCR for estimation of fish abundance and biomass in environmental DNA surveys. PLoS One 10 (3). https://doi.org/10.1371/journal.pone.0122763.

Dong, F., Fang, D., Zhang, Hui, Wei, Q., 2023. Protection and development after the ten-year fishing ban in the Yangtze River. J. Fish. China 42 (2), 1–15. https://doi.org/10.11964/jfc.20221013724.

Emmrich, M., Winfield, I.J., Guillard, J., Rustadbakken, A., Vergès, C., Volta, P., Jeppesen, E., Lauridsen, T.L., Brucet, S., Holmgren, K., Argillier, C., Mehner, T., 2012. Strong correspondence between gillnet catch per unit effort and hydroacoustically derived fish biomass in stratified lakes. Freshw. Biol. 57 (12) https://doi.org/10.1111/fwb.12022.

Feng, C., Wang, K., Xu, W., Yang, L., Wanghe, K., Sun, N., Wu, B., Wu, F., Yang, L., Qiu, Q., Gan, X., Chen, Y., He, S., 2023. Monsoon boosted radiation of the endemic East Asian carps. Sci. China Life Sci. 66 (3) https://doi.org/10.1007/s11427-022-2141-1.

Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. Geogr. Anal. 24 (3) https://doi.org/10.1111/j.1538-4632.1992.tb00261.x.

Guo, X., Wanghe, K., Ahmad, S., Nabi, G., Zhang, K., Zhu, L., Lu, D., Han, D., Zhou, K., Strelnikov, I.I., Khan, T.U., Li, K., Zhao, K., 2022. A methodological framework integrating habitat suitability and landscape connectivity to identify optimal regions for insecticide application: A case study in Tongzhou, China. J. King Saud Univ. Sci. 34 (3) https://doi.org/10.1016/j.jksus.2022.101905.

Hayes, D.B., Bence, J.R., Kwak, T.J., Thompson, B.E., 2007. Abundance, biomass, and production. In: Analysis and Interpretation of Freshwater Fisheries Data, 53. American Fisheries Society, Bethesda, Maryland, pp. 327–374.

Huang, L., Li, J., 2016. Status of Freshwater Fish Biodiversity in the Yangtze River Basin, China. https://doi.org/10.1007/978-981-10-0780-4_2.

Hubert, W.A., Rahel, F.J., 1989. Relations of physical habitat to abundance of four nongame fishes in high-plains streams: a test of habitat suitability index models. N. Am. J. Fish Manag. 9 (3) https://doi.org/10.1577/1548-8675(1989)009<0332:rophta>2.3.co;2.

Jia, Y., Sui, X., Chen, Y., He, D., 2020. Climate change and spatial distribution shaped the life-history traits of schizothoracine fishes on the Tibetan Plateau and its adjacent areas. Glob. Ecol. Conserv. 22 https://doi.org/10.1016/j.gecco.2020.e01041.

Jiang, Z., Jiang, J., Wang, Y., Zhang, E., Zhang, Y., Li, L., Xie, F., Cai, B., Cao, L., Zheng, G., Dong, Lu, 2016. Red list of China's vertebrates. Biodivers. Sci. 24 (5), 500–551. https://doi.org/10.17520/BIODS.2016076.

Kelly, D.J., Hayes, J.W., Allen, C., West, D., Hudson, H., 2015. Evaluating habitat suitability curves for predicting variation in macroinvertebrate biomass with weighted usable area in braided rivers in New Zealand. N. Z. J. Mar. Freshw. Res. 49 (3) https://doi.org/10.1080/00288330.2015.1040424.

Kindong, R., Wu, J., Gao, C., Dai, L., Tian, S., Dai, X., Chen, J., 2020. Seasonal changes in fish diversity, density, biomass, and assemblage alongside environmental variables in the Yangtze River Estuary. Environ. Sci. Pollut. Res. 27 (20) https://doi.org/10.1007/s11356-020-08674-8.

Leslie, P.H., Davis, D.H.S., 1939. An attempt to determine the absolute number of rats on a given area. J. Anim. Ecol. 8 (1) https://doi.org/10.2307/1255.

Li, D., Hao, Y., Duan, Y., 2020. Nonintrusive methods for biomass estimation in aquaculture with emphasis on fish: a review. Rev. Aquac. 12 (3) https://doi.org/10.1111/raq.12388.

Liang, C.E.Z., ChunXin, Z., WenXuan, C., 2016. Evaluating the status of China's continental fish and analyzing their causes of endangerment through the red list assessment. Biodivers. Sci. 24 (5), 598–609.

Mahadevan, G., Bharathirajan, P., Ravi, V., Pouladi, M., Khanghah, M.M.V., 2019. Short communication: age and growth of elongated mudskipper, Pseudapocryptes elongatus (Cuvier, 1816) from Sundarbans, India. Biodiversitas 20 (1). https://doi.org/10.13057/biodiv/d200111.

Murakami, H., Yoon, S., Kasai, A., Minamoto, T., Yamamoto, S., Sakata, M.K., Horiuchi, T., Sawada, H., Kondoh, M., Yamashita, Y., Masuda, R., 2019. Dispersion and degradation of environmental DNA from caged fish in a marine environment. Fish. Sci. 85, 327–337. https://doi.org/10.1007/s12562-018-1282-6.

Radinger, J., Matern, S., Klefoth, T., Wolter, C., Feldhege, F., Monk, C.T., Arlinghaus, R., 2023. Ecosystem-based management outperforms species-focused stocking for enhancing fish populations. Science 379 (6635). https://doi.org/10.1126/science.adf0895.

Rigby, A.S., 2000. Statistical methods in epidemiology. V. Towards an understanding of the kappa coefficient. Disabil. Rehabil. 22 (8) https://doi.org/10.1080/096382800296575.

Říha, M., Prchalová, M., Brabec, M., Draštík, V., Muška, M., Tušer, M., Bartoň, D., Blabolil, P., Čech, M., Frouzová, J., Holubová, M., Jůza, T., R. Moraes, K., Rabaneda-Bueno, R., Sajdlová, Z., Souza, A.T., Šmejkal, M., Vašek, M., Vejřík, L., Kubečka, J., 2023. Calibration of fish biomass estimates from gillnets: step towards broader application of gillnet data. Ecol. Indic. 153, 110425 https://doi.org/10.1016/J.ECOLIND.2023.110425.

Stewart, G., Anderson, R., Wohl, E., 2005. Two-dimensional modelling of habitat suitability as a function of discharge on two Colorado rivers. River Res. Appl. 21 (10) https://doi.org/10.1002/rra.868.

Takahara, T., Minamoto, T., Yamanaka, H., Doi, H., Kawabata, Z., 2012. Estimation of fish biomass using environmental DNA. PLoS One 7 (4). https://doi.org/10.1371/journal.pone.0035868.

Tang, Y., Li, C., Wanghe, K., Feng, C., Tong, C., Tian, F., Zhao, K., 2019. Convergent evolution misled taxonomy in schizothoracine fishes (Cypriniformes: Cyprinidae). Mol. Phylogenet. Evol. 134 (April 2018), 323–337. https://doi.org/10.1016/j.ympev.2019.01.008.

Thuiller, W., Georges, D., Engler, R., 2013. biomod2: Ensemble Platform for Species Distribution Modeling. R Package Version, 2(7).

Tikhonov, G., Opedal, Ø.H., Abrego, N., Lehikoinen, A., de Jonge, M.M.J., Oksanen, J., Ovaskainen, O., 2020. Joint species distribution modelling with the r-package Hmsc. Methods Ecol. Evol. 11 (3) https://doi.org/10.1111/2041-210X.13345.

Wang, H., Wang, P., Xu, C., Sun, Y., Shi, L., Zhou, L., Jeppesen, E., Chen, J., Xie, P., 2022. Can the "10-year fishing ban" rescue biodiversity of the Yangtze River? Innovation 3 (3). https://doi.org/10.1016/j.xinn.2022.100235.

Yin, X., Yang, D., Du, R., 2022. Fishery resource evaluation in Shantou seas based on remote sensing and hydroacoustics. Fishes 7 (4). https://doi.org/10.3390/fishes7040163.

Yue, L., Wang, Y., Zhang, H., Xian, W., 2021. Stock assessment using the LBB method for Portunus trituberculatus collected from the Yangtze Estuary in China. Appl. Sci. 11 (1) https://doi.org/10.3390/app11010342.

Yüksel, F., Demirol, F., Gündüz, F., 2013. Leslie population estimation for Turkish crayfish (Astacus leptodactylus Esch., 1823) in the Keban Dam Lake, Turkey. Turk. J. Fish. Aquat. Sci. 13 (5), 835–839. https://doi.org/10.4194/1303-2712-V13_5_07.

Zandbergen, P.A., 2020. Python Scripting for ArcGIS Pro. Esri Press Redlands, CA, USA. https://www.esri.com/about/newsroom/wp-content/uploads/2020/05/pythonscripting.pdf.

Zhang, C., Zhou, Y., Špoljar, M., Fressl, J., Tomljanović, T., Rama, V., Kuczyńska-Kippen, N., 2023. How can top-down and bottom-up manipulation be used to mitigate eutrophication? Mesocosm experiment driven modeling zooplankton seasonal dynamic approach in the trophic cascade. Water Res. 243 https://doi.org/10.1016/j.watres.2023.120364.

Zhang, H., Kang, M., Shen, L., Wu, J., Li, J., Du, H., Wang, C., Yang, H., Zhou, Q., Liu, Z., Gorfine, H., Wei, Q., 2020. Rapid change in Yangtze fisheries and its implications for global freshwater ecosystem management. Fish Fish. 21 (3) https://doi.org/10.1111/faf.12449.

Zhao, K., Duan, Z.Y., Peng, Z.G., Guo, S.C., Li, J.B., He, S.P., Zhao, X.Q., 2009. The youngest split in sympatric schizothoracine fish (Cyprinidae) is shaped by ecological adaptations in a Tibetan Plateau glacier lake. Mol. Ecol. 18 (17), 3616–3628. https://doi.org/10.1111/j.1365-294X.2009.04274.x.

Zhu, R., He, D., Feng, X., Xiong, W., Tao, J., 2021. The new record of the highest distribution altitude of cyprinid fishes in the world. J. Appl. Ichthyol. 37 (3) https://doi.org/10.1111/jai.14204.