



OPEN The species distribution model based on the random forest algorithm reveals the distribution patterns of *Neophocaena asiaeorientalis*

Rao Rongcheng^{1,2}, Huang Yi^{1,2}, Min Jialing^{1,2}, Yang Ying^{1,2}, Liu Fanning^{1,2}, Wu Xiya^{1,2}, Shi Xinyuan^{1,2}, Li Caigang^{1,2}, Dai Yingen^{1,2}, Huang Qinghai^{1,2}, Yu Jinxiang^{1,2}✉ & Que Jianglong^{1,2}✉

The Species Distribution Model (SDM) provides a crucial foundation for the conservation of the Yangtze finless porpoise (YFP), a critically endangered freshwater cetacean endemic to China. In this study, we conducted population and habitat surveys, and employed the Random Forest algorithm (RF) to construct SDMs. We found that the habitat preference of YFP shows complex seasonality. Cyanobacteria and total phosphates have been identified as the predominant factors influencing the YFP distributions by affecting prey resources. We emphasize that ascertaining the presence and pseudo-absence points of YFP, in conjunction with the selection of key factors, constitutes the foundational element in the construction of SDMs. We suggest that the incorporation of techniques such as environmental DNA could expand the range of environmental factors, particularly with regard to the distribution of prey resources at the genus or species level. This study provides guidance for the SDMs of YFP and demonstrates the potential of machine learning algorithms in constructing SDMs for the endangered aquatic species.

Keywords Conservation biology, Endangered aquatic species, Habitat preference, Hydrographic environment, Machine learning algorithm

Habitat plays a pivotal role in the distribution and abundance of species and is the focus of conservation efforts^{1,2}. The Species Distribution Models (SDM) have become effective tools for species and habitat conservation by integrating species abundance observations with environmental estimates. They can predict potentially suitable habitat for species³, identify high-priority areas for monitoring and conservation^{4–6}, and assess the influences of environmental changes on species⁷. The SDMs are constrained by the ecological principles and assumptions in theory and application⁸. In particular, the influence of environment on species is complex⁹. It is necessary to select primary limiting and determining variables for SDMs¹⁰. The machine learning algorithms, which learn patterns from extensive data to perform prediction and classification, have become an effective method for constructing SDMs¹¹. Among these, the Random Forest algorithm (RF) not only performs extremely well in ecological predictions but also identifies key factors¹². With the increasing availability of big-data, the use of RF in SDMs has become more prevalent^{13,14}. The environmental layers used to construct SDMs base on RF are typically derived from remote sensing satellites^{15,16}. The aquatic factors such as dissolved oxygen, which are vital for aquatic organisms¹⁷, cannot be acquired through remote sensing¹⁸. As a result, the application of RF is challenging in SDMs for endangered aquatic species^{19,20}. Empirical research on the distribution and habitat utilization of endangered aquatic species is needed urgently to provide references.

The Yangtze finless porpoise (YFP, *Neophocaena asiaeorientalis*) is a critically endangered freshwater cetacean that inhabits the middle-lower Yangtze channel and two large appended lakes, namely Poyang and Dongting²¹. Poyang Lake is one of the most critical habitats, supporting almost half of the YFP populations²². Field surveys indicate YFPs primarily distribute along the main channel between Hukou and Poyang, and also

¹Aquatic Conservation and Rescue Center of Jiangxi Province, Nanchang 330096, China. ²Poyang Lake Yangtze Finless Porpoise Conservation Base of Ministry of Agriculture and Rural Affairs, Nanchang 330096, China. ✉email: 7835711@qq.com; que_jianglong@sina.com

in some large sandpit areas between Duchang and Yongxiu^{23,24}. These field surveys do not explain how the YFP preference habitats in Poyang Lake as well as the driving forces. Previous study found that the YFP sighting rates only showed a negative correlation with the density of ships in the main channel^{25,26}. However, due to the hydrological condition varies in Poyang Lake, the YFP population distributions showed drastic fluctuation seasonally²⁴. It seemed that hydrological conditions also played a considerable role in their habitat preference. In Poyang Lake, increasingly frequent human activities, such as heavy shipping and sand mining, have severely squeezed their living space, almost all the YFP populations suffer from habitat loss²⁷. Therefore, studies on SDMs for YFP become extremely important, as they help delineate suitable areas for targeted conservation efforts.

In this study, we conducted habitats surveys in Poyang Lake, investigating the distribution of the YFP populations and the aquatic factors of habitats. The result were used to construct SDMs base on RF. We analyzed the impact of aquatic factors on the YFP distribution patterns. The objective of this study was to elucidated how YFPs select their habitats and to provide information for the management and conservation of both populations and habitats. Based on above analysis, we also discussed perspectives on the SDM for YFPs. Our research also demonstrated the applicability of machine learning algorithms in SDMs for endangered aquatic species.

Materials and methods

Study area and monitoring point

The study area is a large sandpit area in Poyang Lake (see Supplementary Fig.S1), with a total water area of approximately 20 km² and minimal evidence of human activity^{28,29}. In accordance with the technical specifications for surface water environmental quality monitoring issued by the Chinese Ministry of Agriculture (HJ 91.2—2022), a total of 12 sampling points (p1-p12 in Fig. 1) were established. Four surveys were conducted in April, June, September, and December 2023.

Environmental variables

The aquatic factors were collected from sampling points in accordance with the technical specifications for surface water environmental quality monitoring (HJ 91.2—2022) and the technical guidelines for water ecological monitoring-aquatic organism monitoring and evaluation of lakes and reservoirs (HJ 1296—2023), both issued by the Chinese Ministry of Agriculture. The data were subsequently imported into ArcGIS (version 10.8). The inverse distance weighting (IDW) method was employed to generate the aquatic factor layers.

The 57 aquatic factor layers (25 m x 25 m) were created in a GIS format as proxies for habitat predictors (see Supplementary Table.S1). These layers included water physical and chemical indicators, the comprehensive trophic level index (*TLI*)³⁰, the aquatic biology factors (the density and biomass of zooplankton, phytoplankton and benthic), and the biodiversity indicators (the Shannon-Wiener index (*H*), the Simpson index (*D*), the Margalef index (*M*) and the Pielou index (*J*) of zooplankton, phytoplankton and benthic).

Presence and pseudo-absence points of YFPs

A research vessel was operated at a speed of approximately 12 km/h in the study area. Five observers were positioned at the bow of the vessel to observe YFPs during the voyage. The latitude and longitude coordinates of the YFP occurrences were recorded with the hand-held GPS (GPSMAP639CSX, Garmin, America)^{22,27}. An acoustic event recorder (RPCD-A2, PinDu, China) was installed at the stern of the vessel to assist visual observations³¹.

The study area was divided into equal-sized grids (0.1 × 0.1 km) and the geographic coordinates of the center point of grids were extracted in ArcGIS 10.8. Both the geographic coordinates of the grids and YFPs were imported into RStudio, where the R package “geosphere” was employed to calculate the minimum spatial distance (MSD) between grids and YFPs. The points with the MSD less than the thresholds (the threshold was set from 100 m to 1000 m) were classified as presence points, while the remaining points were classified as pseudo-absence points.

Construction of SDMs

The 57 aquatic factor layers were imported into RStudio, and the R package “abundanceR” was employed to extract aquatic factor data of the presence and pseudo-absence points. The SDMs were created base on RF with the R package “randomForest” in accordance with default parameters. The results of each survey were used to construct independent SDMs, a total of four initial SDMs were obtained.

Lists of factors ranked by RF in order of feature importance (mean decrease accuracy index and mean decrease Gini index) were determined over 100 iterations. The number of key factors was identified using 10-fold cross-validation implemented with the `rfcv()` function in the R package “randomForest”. When the cross-validation error was minimum, the top-ranked importance factors across multiple SDMs were selected to construct the reconstructed SDMs.

The performance of SDMs

The SDMs were used to make cross-predictions the presence and pseudo-absence area of YFPs with the `predict()` function in the R package “stat”. The predicted results were then used to calculate the receiver operating characteristic (ROC) curves, the True Skill Statistic (TSS) and the area under the ROC curve (AUC) in the R package “pROC” and “tidysdm”.

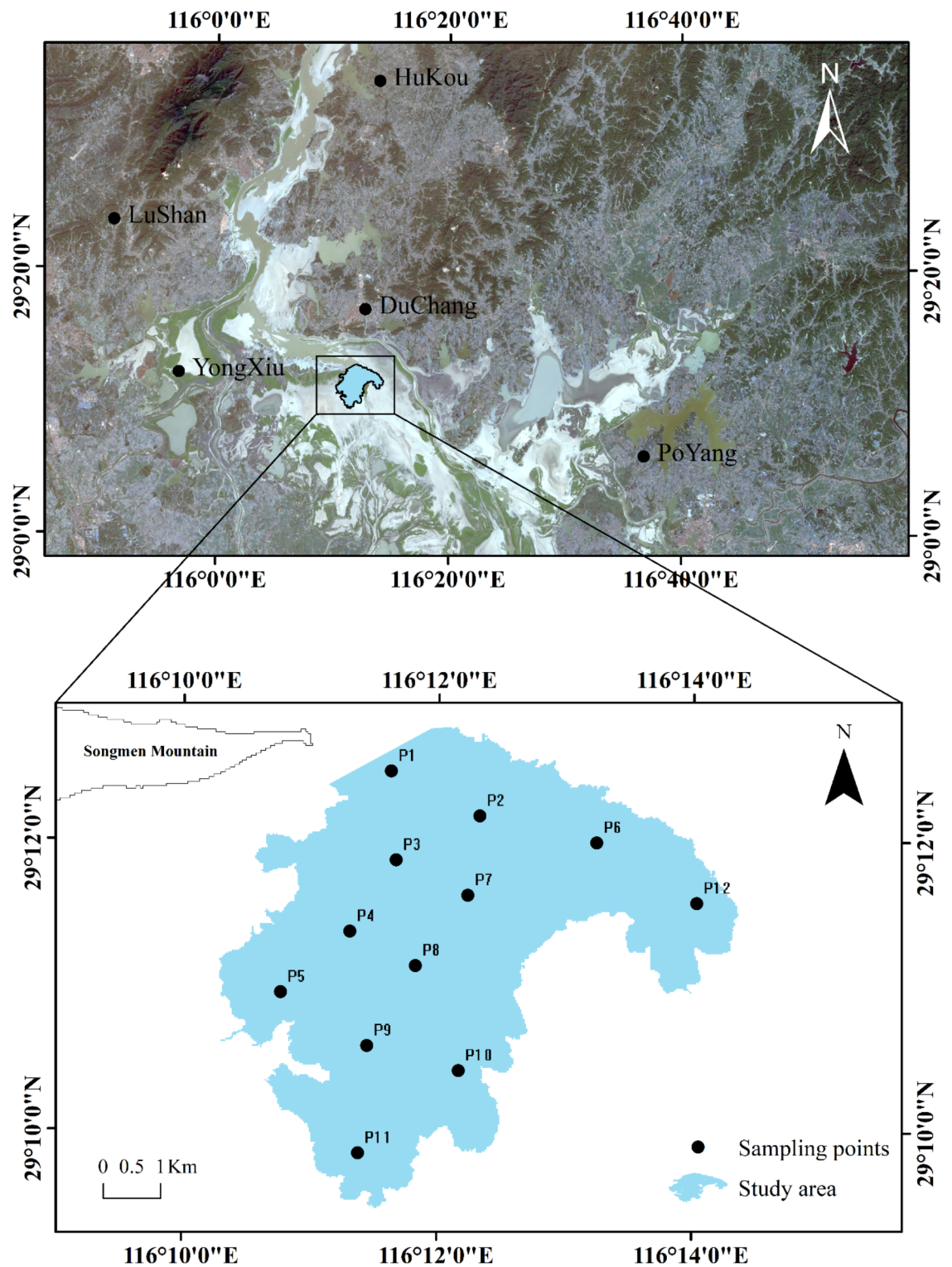


Fig. 1. Study area and sampling points. (It was generated by ArcMap 10.8 software (<https://www.esri.com/en-us/arcgis/products/arcgis-desktop/overview>)).

Results

The results of habitat environments survey

The four surveys identified a total of 26, 38, 37 and 21 groups of YFPs. A few of YFPs were found outside the study area due to rising water levels during the rainy season (Fig. 2).

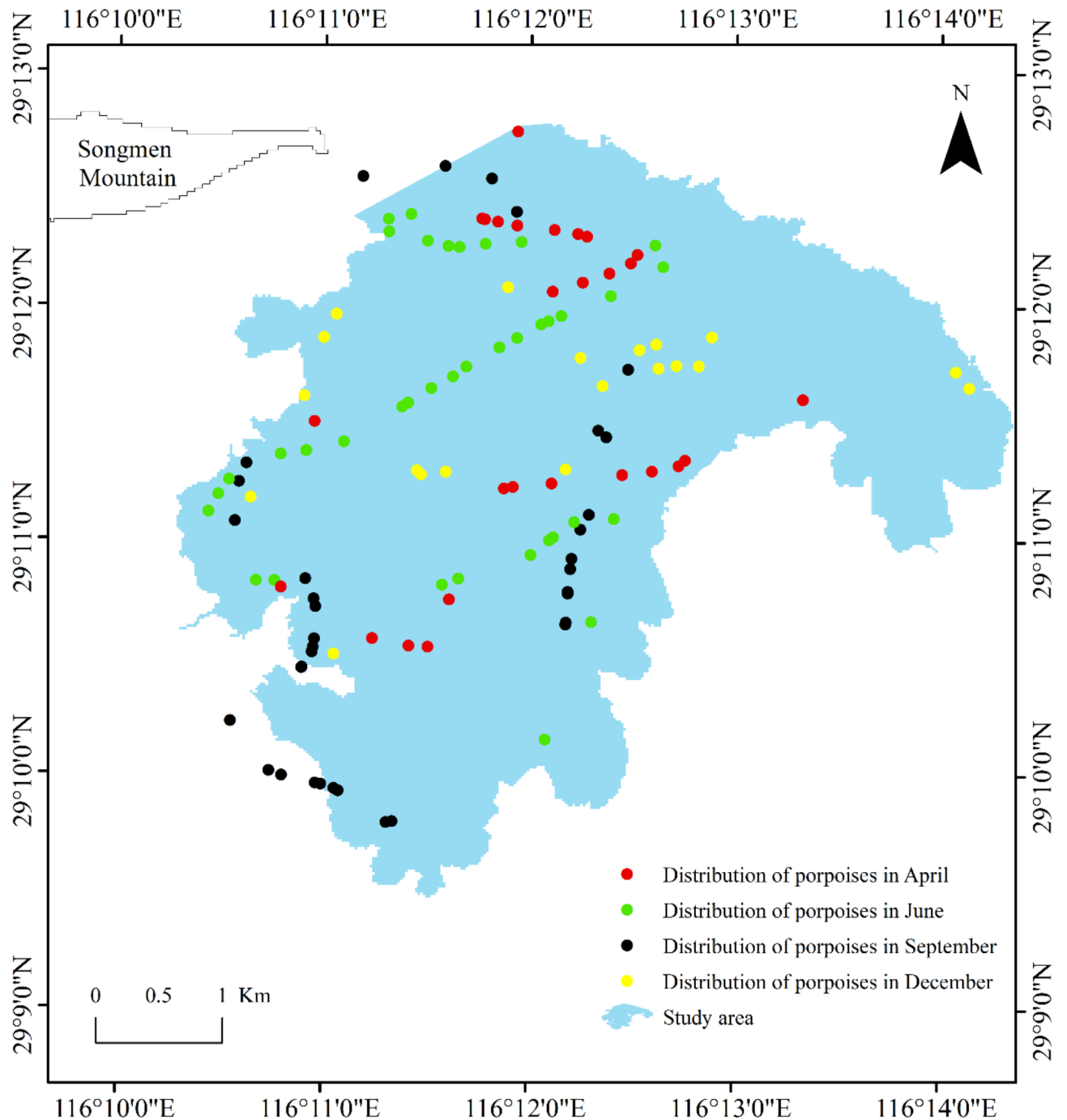


Fig. 2. Distribution of YFPs in each survey.

The most appropriate thresholds

As the thresholds increased, the out-of-bag (OOB) estimate error rate and the classification error rate of the presence points progressively decreased. The classification error rate of the pseudo-absence points increases. These three curves intersected when the threshold reaches approximately 550 m (Fig. 3).

The use of SDMs

The SDMs exhibited accurate predictions of the YFP distributions for the same period, with all YFPs were distributed in the areas predicted by the SDMs (Fig. 4 and Supplementary Fig.S2-4). The SDMs exhibited limited efficacy in predicting the YFP distributions for other survey periods, with only a modest number of YFPs distributed in the areas predicted by SDMs (Fig. 4 and Supplementary Fig.S2-4); In some instances, the prediction offered by SDMs are particularly extreme. They assert that the entirety of the study area is either a presence area or an absence area (Supplementary Fig.S2-4).

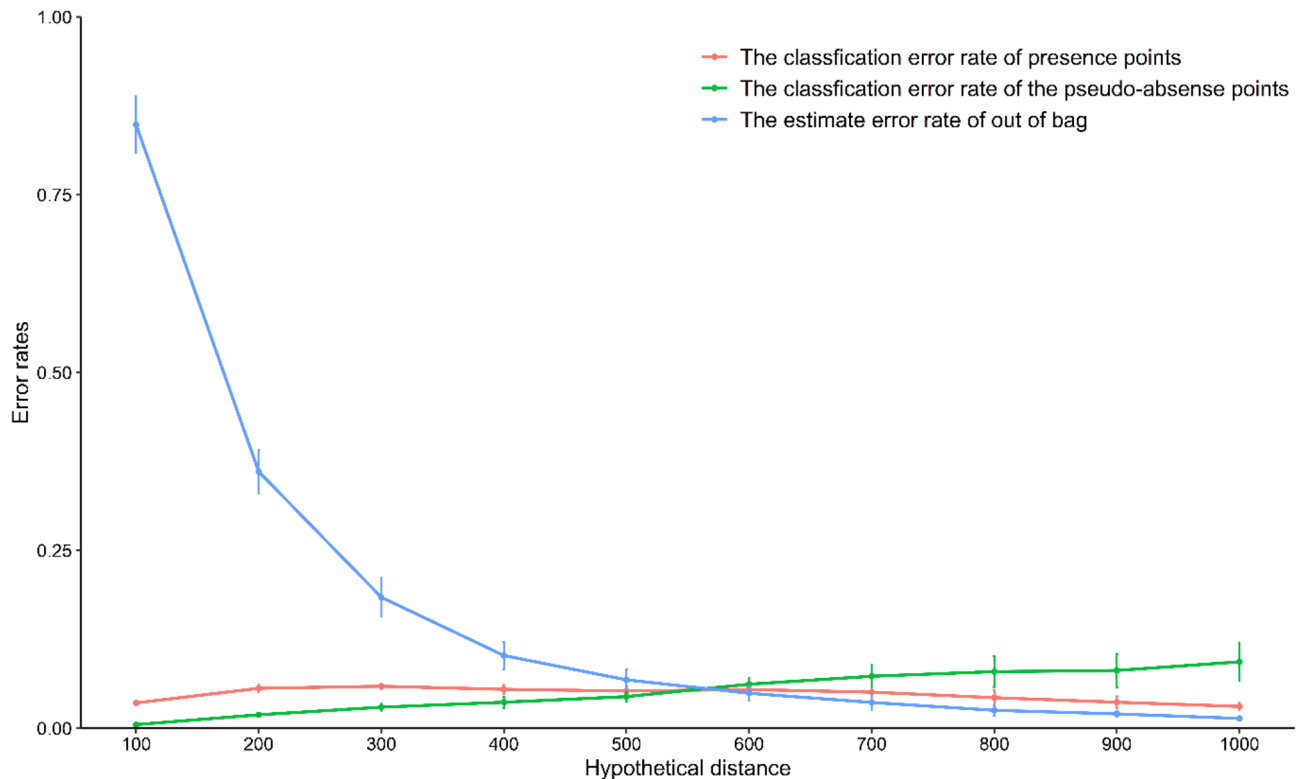


Fig. 3. Relationship between the SDMs accuracy and thresholds.

The contribution of aquatic factors to SDMs

The cross-validation error curve stabilized when 15 factors were used in the SDMs and these top-ranked aquatic factors in each SDMs are not identical. The 13 aquatic factors appeared in multiple top aquatic factors lists of SDMs (Fig. 5 and Supplementary Table.S2). The importance of the aquatic factors for SDMs is visualized in Supplementary Fig.S5.

The performance of SDMs

The AUC and TSS of the reconstructed SDMs (excluding September) were higher than those of the initial SDMs (Fig. 6). The ROC curves of the initial and reconstructed SDMs are displayed in Supplementary Fig.S6-9.

Discussion

The population distributions

Previous studies found that YFPs migrate between Poyang Lake and the mainstream of the Yangtze River in response to hydrological variations³². We found that the abundance and distribution of YFPs were not exactly the same in each survey (Fig. 2), which suggests that the migration of YFPs occurs between the sandpit and the main channel. This also demonstrates the YFP populations have seasonal habitat preferences²⁸. These hydrological influences not only influenced on the YFP distributions across the whole-lake scale, but also work at fine-scales.

The principal limiting factors

In this study, the prediction maps were not entirely accurate in delineating the actual presence of YFPs (Fig. 4). This limitation stems from SDMs predicting the Suitability of presence for YFP, reflecting constraints of binary classification in habitat prediction³³. Suitability distributions are theoretical representations of the potential distribution of species. However, specie distributions do not necessarily align with the most suitable habitats unless the population reaches its carrying capacity². Consequently, it is imperative to prioritize the consideration of the aquatic factors, which play a major role in the SDMs^{34,35}, rather than the prediction of SDMs.

We conducted comprehensive habitat surveys, including water level and water flow measurements, which correlate with YFP distribution^{36,37}. The aquatic factors that had not been validated were quantified, including dissolved oxygen and diversity of aquatic organisms³⁸. The numerous and multifaceted aquatic factors ensured the accuracy of SDMs. This also significantly influenced the model transferability, as evidenced by the performance of the initial SDMs in predicting other periods ($AUC = 0.557 \pm 0.062$, $TSS = 0.18 \pm 0.062$). The reconstructed SDMs using top-ranked factors showed improved performance ($AUC = 0.624 \pm 0.056$, $TSS = 0.30 \pm 0.11$). AUC and TSS is shown to be independent of prevalence³⁹. The performance of the models is evaluated by AUC and TSS based on a dichotomous prediction of presence-absence in this study. This results in suboptimal SDM performance, and an $AUC > 0.75$ is generally considered a good result. However, our study underscores the importance of selecting key factors for constructing SDM of YFP.

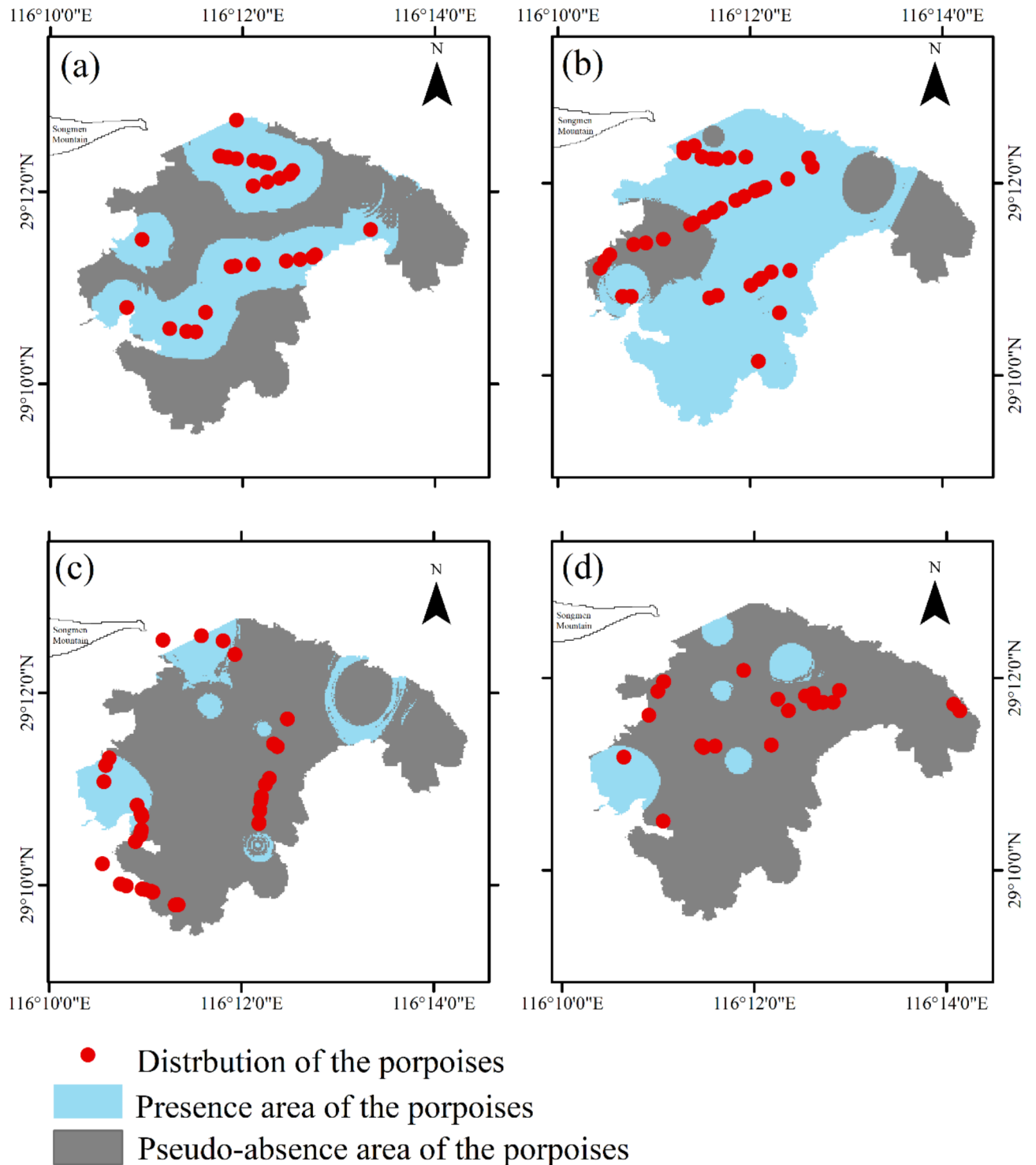


Fig. 4. Prediction maps of the SDM that based on the April habitat survey. (a) The SDM predictions for the April survey. (b) The SDM predictions for the June survey. (c) The SDM predictions for the September Survey. (d) The SDM predictions for the December Survey.

There were no factors appeared in the lists of top-ranked importance factors for all SDMs. This suggests seasonally variable habitat preferences in YFPs, and the seasons cannot be simply categorized as rainy or dry seasons. We also found that total phosphate concentration and cyanobacteria density, appeared in the list of top-ranked importance factors for three SDMs (April, June and September). Previous research found that one of the most important factors affecting the distribution of YFPs is prey resources⁴⁰. It is speculated that these factors may have indirectly influenced the distribution of YFPs by affecting prey resources.

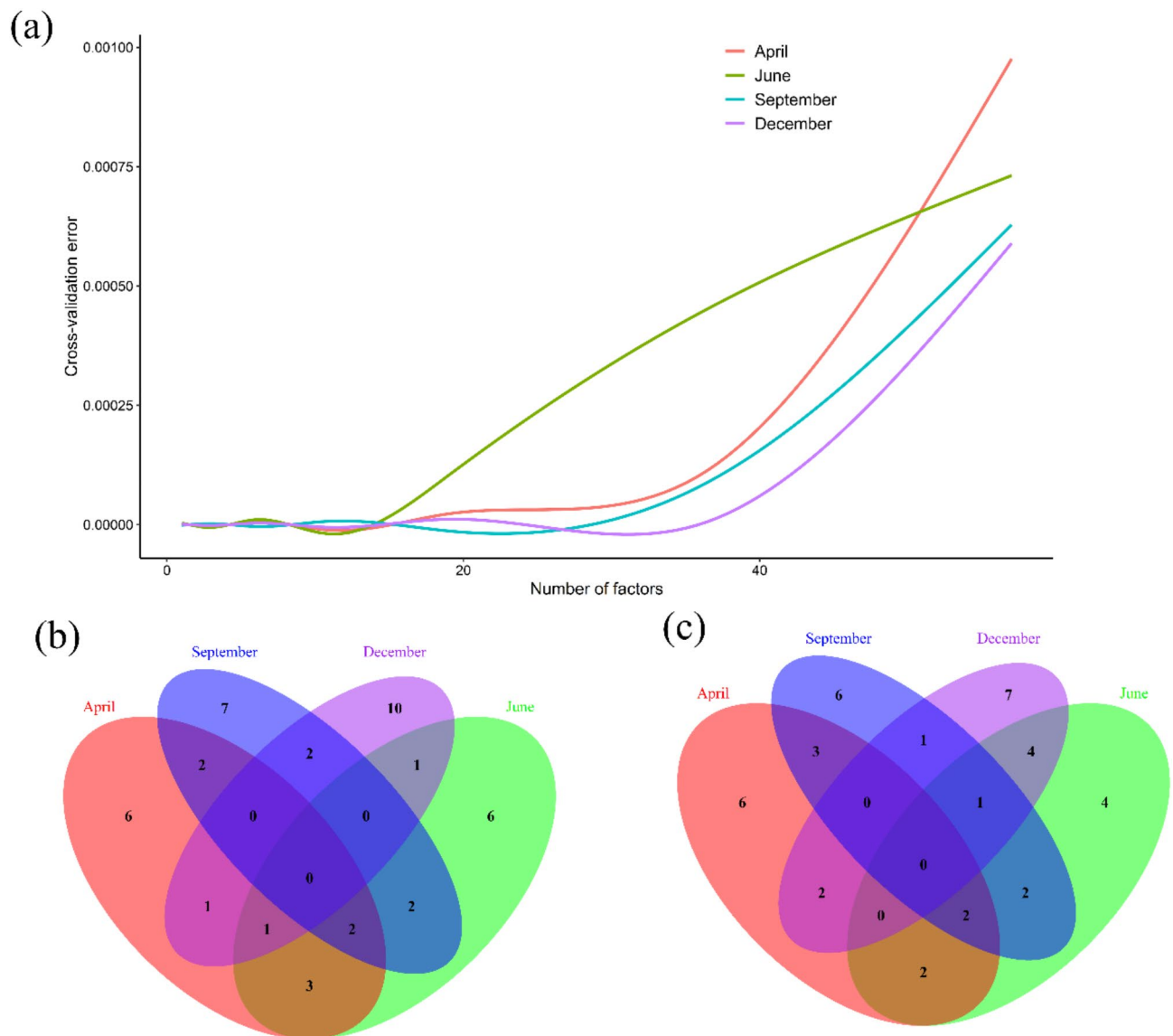


Fig. 5. Degree of contribution of aquatic factors to the SDMs. (a) The 15 top importance aquatic factors were identified by RF in the four SDMs. (b) The occurrence of top aquatic factors based on mean decrease accuracy index in the multiple SDMs. (c) The occurrence of top aquatic factors based on mean decrease Gini index in the multiple SDMs.

The perspectives on SDMs for YFP

The objective of field surveys is to ascertain the presence points of species. However, the absence points are also crucial for SDMs. Consequently, researchers use pseudo-absences in place of true absences, which are established under multiple assumptions with careful planning⁴¹. Due to the restricted distribution range and high mobility of YFPs, a series of thresholds were set for the MSD⁴². We found that the threshold was most appropriate when set at approximately 550 m in this study, all four SDMs performed well under this threshold. It is speculated that YFPs occupy the surrounding 550 m area. But the behavioral characteristics of YFP are extremely complex⁴³. Therefore, it is necessary to adopt a suitable approach for determining the threshold for each SDM, e.g., by observing the model performs across different thresholds.

In this study, the study area is a natural habitat with little or no human disturbance. This establishes a foundation for the acquisition of the baseline of natural habitat preferences³⁸. However, inevitable overlap between the activity space of YFPs and that of humans in most habitats²⁶. Research showed that factors such as ship noise have been demonstrated to exert an influence on the natural behavior of the YFPs⁴⁴. It is suggested that anthropogenic intensity indicators need to be included in SDMs for YFPs.

Additionally, prey resources were not studied due to no-fishing policy constraints. Cetaceans are highly selective in habitat utilization which is influenced by prey resources⁴⁵. The diet of YFPs consists only of specific small fish⁴⁶. Although acoustic equipment can monitor prey populations, it cannot distinguish species⁴⁷. This posed challenge for incorporation prey resources into SDMs. In recent years, environmental DNA (eDNA)

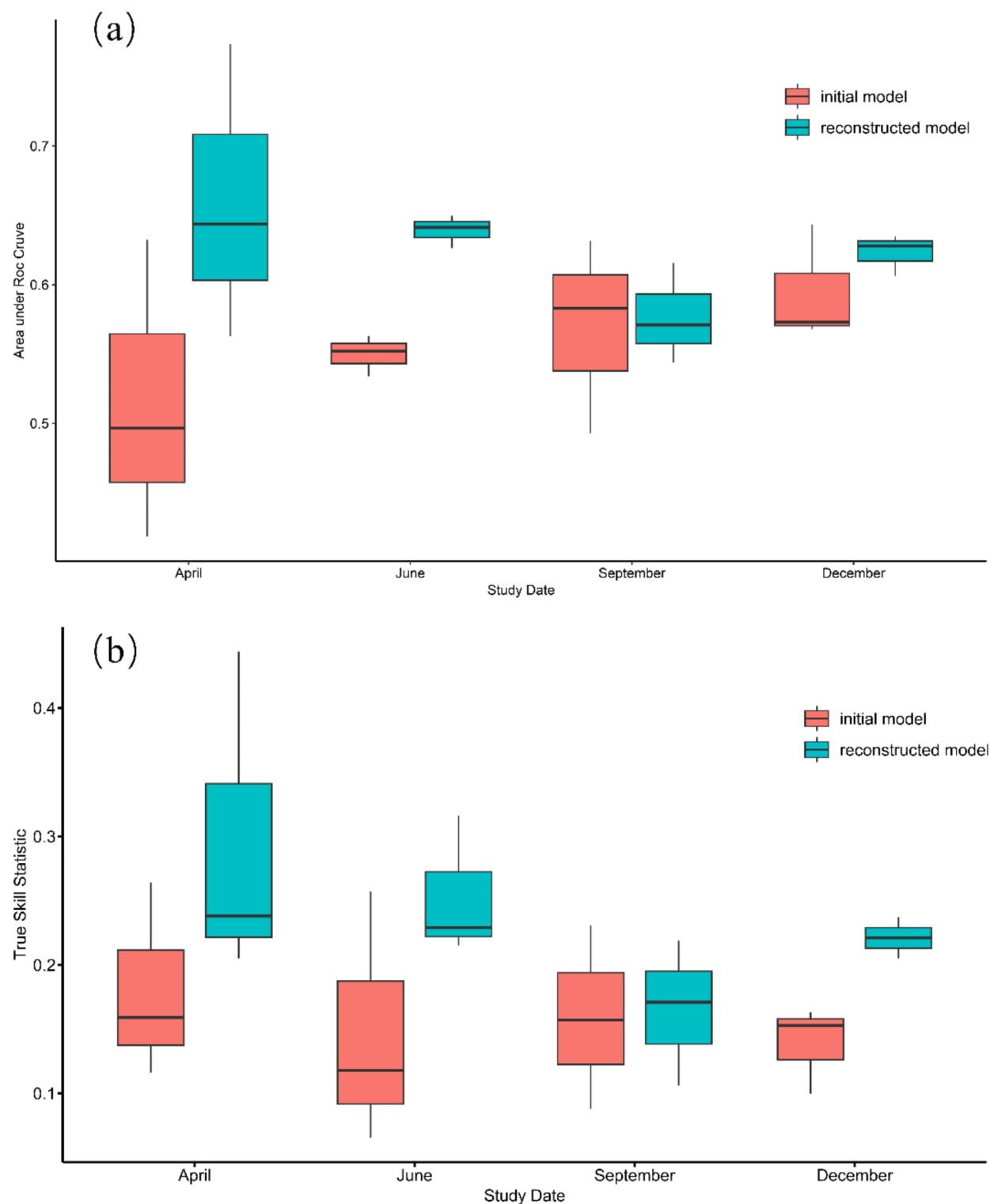


Fig. 6. (a) The AUC of the initial and reconstructed SDMs. (b) The TSS of the initial and reconstructed SDMs.

technology has matured and demonstrated particular efficacy in analyzing fish community structure⁴⁸. The integration of eDNA with acoustic monitoring is anticipated to address this challenge.

Conclusion

In this study, we employed RF for the first time to construct SDMs for YFPs and investigate the principal limiting factors affecting the YFP distributions. We found that YFPs were affected by different environmental factors in different seasons, total phosphorus concentration and cyanobacteria density were identified as common key

factors. Our findings provide crucial insights to inform the conservation management of YFPs and serve as a theoretical foundation for the developing SDMs for endangered aquatic species.

Data availability

All data generated or analysed during this study are available from the corresponding author on request.

Received: 20 August 2024; Accepted: 27 February 2025

Published online: 23 March 2025

References

- Andrewartha, H. G. & Birch, L. C. The distribution and abundance of animals. *Sci. (New York N Y)*. **121**, 389–390. <https://doi.org/10.1126/science.121.3142.389> (1955).
- Boyce, M. S. et al. Can habitat selection predict abundance? *J. Anim. Ecol.* **85**, REVIEW, 11–20. <https://doi.org/10.1111/1365-2656.12359> (2016).
- Wei, J., Niu, M., Zhang, H., Cai, B. & Ji, W. Global potential distribution of invasive species *Pseudococcus viburni* (Hemiptera: Pseudococcidae) under climate change. *Insects* **15**, 195. <https://doi.org/10.3390/insects15030195> (2024).
- Kafash, A., Ashrafi, S. & Yousefi, M. Modeling habitat suitability of bats to identify high priority areas for field monitoring and conservation. *Environ. Sci. Pollut. Res. Int.* **29**, 25881–25891. <https://doi.org/10.1007/s11356-021-17412-7> (2022).
- Bosso, L. et al. Integrating citizen science and Spatial ecology to inform management and conservation of the Italian seahorses. *Ecol. Inf.* **79**, 102402. <https://doi.org/10.1016/j.ecoinf.2023.102402> (2024).
- Rêgo, M., Del-Rio, G. & Brumfield, R. Subspecies-level distribution maps for birds of the Amazon basin and adjacent areas. *J. Biogeogr.* **51**. <https://doi.org/10.1111/jbi.14718> (2023).
- Ahmed, A. S., Bekele, A., Kasso, M. & Atickem, A. Impact of climate change on the distribution and predicted habitat suitability of two fruit bats (*Rousettus aegyptiacus* and *Epomophorus labiatus*) in Ethiopia: implications for conservation. *Ecol. Evol.* **13**, e10481. <https://doi.org/10.1002/ece3.10481> (2023).
- Guisan, A. & Thuiller, W. Predicting species distribution: Offering more than simple habitat models. *Ecol. Lett.* **8**, 993–1009. <https://doi.org/10.1111/j.1461-0248.2005.00792.x> (2005).
- Austin, M. P. Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. *Ecol. Model.* **157**, 101–118. [https://doi.org/10.1016/S0304-3800\(02\)00205-3](https://doi.org/10.1016/S0304-3800(02)00205-3) (2002).
- Pinto-Ledezma, J. N. & Cavender-Bares, J. Predicting species distributions and community composition using satellite remote sensing predictors. *Sci. Rep.* **11**, 16448–16460. <https://doi.org/10.1038/s41598-021-96047-7> (2021).
- Jiang, T., Gradus, J. L. & Rosellini, A. J. Supervised machine learning: A brief primer. *Behav. Ther.* **51**, 675–687. <https://doi.org/10.1016/j.beth.2020.05.002> (2020).
- Mi, C., Huettmann, F., Guo, Y., Han, X. & Wen, L. Why choose random forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ* **5**, e2849. <https://doi.org/10.7717/peerj.2849> (2017).
- Huettmann, F. et al. A super SDM (species distribution model) 'in the cloud' for better habitat-association inference with a 'big data' application of the great Gray Owl for Alaska. *Sci. Rep.* **14**, 7213. <https://doi.org/10.1038/s41598-024-57588-9> (2024).
- Steiner, M., Huettmann, F., Bryans, N. & Barker, B. With super SDMs (machine learning, open access big data, and the cloud) towards more holistic global squirrel hotspots and coldspots. *Sci. Rep.* **14**, 5204. <https://doi.org/10.1038/s41598-024-55173-8> (2024).
- Meichenich, M. F. & Žliobaitė, I. Eco-ISEA3H, a machine learning ready Spatial database for ecometric and species distribution modeling. *Sci. Data*. **10**, 77. <https://doi.org/10.1038/s41597-023-01966-x> (2023).
- Bonannella, C. et al. Forest tree species distribution for Europe 2000–2020: mapping potential and realized distributions using Spatiotemporal machine learning. *PeerJ* **10**, e13728. <https://doi.org/10.7717/peerj.13728> (2022).
- Aarts, B. G. W., Van Den Brink, F. W. B. & Nienhuis, P. H. Habitat loss as the main cause of the slow recovery of fish faunas of regulated large rivers in Europe: the transversal floodplain gradient. *River Res. Appl.* **20**, 3–23. <https://doi.org/10.1002/rra.720> (2004).
- Wasehun, E. T., Beni, H., Di Vittorio, C. A. & L. & UAV and satellite remote sensing for inland water quality assessments: A literature review. *Environ. Monit. Assess.* **196**, 277. <https://doi.org/10.1007/s10661-024-12342-6> (2024).
- Lee, Z., Churnside, J., Mao, Z., Wu, S. & Zibordi, G. Active and passive optical remote sensing of the aquatic environment: Introduction to the feature issue. *Appl. Opt.* **59**, Aps1–aps2. <https://doi.org/10.1364/ao.392549> (2020).
- Elith, J. & Leathwick, J. R. Species distribution models: Ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Syst.* **40**, 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159> (2009).
- Zhao, X. et al. Abundance and conservation status of the Yangtze finless porpoise in the Yangtze river, China. *Biol. Conserv.* **141**, 3006–3018. <https://doi.org/10.1016/j.biocon.2008.09.005> (2008).
- Huang, J. et al. Population survey showing hope for population recovery of the critically endangered Yangtze finless porpoise. *Biol. Conserv.* **241**, 108315. <https://doi.org/10.1016/j.biocon.2019.108315> (2020).
- Chen, M. et al. Parentage-Based group composition and dispersal pattern studies of the Yangtze finless porpoise population in Poyang lake. *Int. J. Mol. Sci.* **17**. <https://doi.org/10.3390/ijms17081268> (2016).
- Jiang-Long, Q. et al. Population and distribution characteristics of Yangtze finless porpoise in Jiangxi waters during the dry season. *Acta Hydrobiol. Sin.* **47**, 1701–1708. <https://doi.org/10.7541/2023.2022.0503> (2023).
- Liu, X. et al. Seasonal Yangtze finless porpoise (*Neophocaena asiaeorientalis asiaeorientalis*) movements in the Poyang lake, China: Implications on flexible management for aquatic animals in fluctuating freshwater ecosystems. *Sci. Total Environ.* **807**, 150782. <https://doi.org/10.1016/j.scitotenv.2021.150782> (2022).
- Mei, Z. et al. Mitigating the effect of shipping on freshwater cetaceans: The case study of the Yangtze finless porpoise. *Biol. Conserv.* **257**, 109132. <https://doi.org/10.1016/j.biocon.2021.109132> (2021).
- Huang, S. L. et al. Saving the Yangtze finless porpoise: Time is rapidly running out. *Biol. Conserv.* **210**, 40–46. <https://doi.org/10.1016/j.biocon.2016.05.021> (2017).
- Rongcheng, R. et al. Population structure and growth status of Yangtze finless porpoise in the sand pits waters of Southern Songmen mountain, Poyang lake. *J. Shanghai Ocean. Univ.* **32**, 1237–1244. <https://doi.org/10.12024/jso> (2023).
- Rongcheng, R. et al. Analysis of the population genetics of the Yangtze finless porpoise population in the Poyang lake basin: Insight of conservation recommendations. **34**, e4145. (2024). <https://doi.org/10.1002/aqc.4145>
- Leng, M. et al. Assessment of Water Eutrophication at Bao'an Lake in the Middle Reaches of the Yangtze River Based on Multiple Methods. *Int. J. Environ. Res. Public Health*. **20**. <https://doi.org/10.3390/ijerph20054615> (2023).
- Li, W. et al. A real-time passive acoustic monitoring system to detect Yangtze finless porpoise clicks in Ganjiang river, China. *Front. Mar. Sci.* **9**. <https://doi.org/10.3389/fmars.2022.883774> (2022).

32. Duan, P. X. et al. Anthropogenic activity, hydrological regime, and light level jointly influence Temporal patterns in biosonar activity of the Yangtze finless porpoise at the junction of the Yangtze river and Poyang lake, China. *Zoological Res.* **44**, 919–931. <https://doi.org/10.24272/j.issn.2095-8137.2022.504> (2023).
33. Guisan, A. & Zimmermann, N. E. Predictive habitat distribution models in ecology. *Ecol. Model.* **135**, 147–186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9) (2000).
34. Bosso, L. et al. The rise and fall of an Alien: why the successful colonizer *Littorina saxatilis* failed to invade the mediterranean sea. *Biol. Invasions.* **24**, 3169–3187. <https://doi.org/10.1007/s10530-022-02838-y> (2022).
35. Hu, Z. M. et al. Intraspecific genetic variation matters when predicting seagrass distribution under climate change. *Mol. Ecol.* **30**, 3840–3855. <https://doi.org/10.1111/mec.15996> (2021).
36. Wang, Z. T., Duan, P. X., Akamatsu, T., Wang, K. X. & Wang, D. Temporal and Spatial biosonar activity of the recently established uppermost Yangtze finless porpoise population downstream of the Gezhouba dam: correlation with hydropower cascade development, shipping, hydrological regime, and light intensity. *Ecol. Evol.* **14**, e11346. <https://doi.org/10.1002/ece3.11346> (2024).
37. Li, Q. et al. Habitat configuration of the Yangtze finless porpoise in Poyang lake under a shifting hydrological regime. *Sci. Total Environ.* **838**, 155954. <https://doi.org/10.1016/j.scitotenv.2022.155954> (2022).
38. Mei, Z. et al. Habitat preference of the Yangtze finless porpoise in a minimally disturbed environment. *Ecol. Model.* **353**, 47–53. <https://doi.org/10.1016/j.ecolmodel.2016.12.020> (2017).
39. Manel, S., Williams, H. C. & Ormerod, S. J. Evaluating presence-absence models in ecology: The need to account for prevalence. *J. Appl. Ecol.* **38**, 921–931. <https://doi.org/10.1046/j.1365-2664.2001.00647.x> (2001).
40. Zhang, X. et al. Effects of fish community on occurrences of Yangtze finless porpoise in confluence of the Yangtze and Wanhe rivers. *Environ. Sci. Pollut. Res.* **22**, 9524–9533. <https://doi.org/10.1007/s11356-015-4102-x> (2015).
41. Elith, J. et al. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129–151. <https://doi.org/10.1111/j.2006.0906-7590.04596.x> (2006).
42. Barbet-Massin, M., Jiguet, F., Albert, C. H. & Thuiller, W. Selecting pseudo-absences for species distribution models: how, where and how many? **3**, 327–338, (2012). <https://doi.org/10.1111/j.2041-210X.2011.00172.x>
43. Wei, Z. et al. Population size, behavior, movement pattern and protection of Yangtze finless porpoise at Balijiang section of the Yangtze river. *Resour. Environ. Yangtze Basin.* **11**, 427–432 (2002).
44. Wang, Z. T. et al. Underwater noise pollution in China's Yangtze river critically endangers Yangtze finless porpoises (*Neophocaena asiaeorientalis asiaeorientalis*). *Environ. Pollut.* **262**, 114310. <https://doi.org/10.1016/j.envpol.2020.114310> (2020).
45. Brough, T. E., Rayment, W. J., Slooten, L. & Dawson, S. Prey and habitat characteristics contribute to hotspots of distribution for an endangered coastal Dolphin. **10**, (2023). <https://doi.org/10.3389/fmars.2023.1204943>
46. Wang, Z., Akamatsu, T., Wang, K. & Wang, D. The diel rhythms of biosonar behavior in the Yangtze finless porpoise (*Neophocaena asiaeorientalis asiaeorientalis*) in the Port of the Yangtze river: the correlation between prey availability and boat traffic. *PLoS One.* **9**, e97907. <https://doi.org/10.1371/journal.pone.0097907> (2014).
47. Crossin, G. T. et al. Acoustic telemetry and fisheries management. *Ecol. Appl.* **27**, 1031–1049. <https://doi.org/10.1002/eap.1533> (2017).
48. Minamoto, T. & Environmental DNA analysis for macro-organisms: Species distribution and more. *DNA Res.* **29** <https://doi.org/10.1093/dnares/dsac018> (2022).

Acknowledgements

The authors are grateful to the anonymous reviewers for their constructive suggestions and comments that have helped to improve the quality of this paper.

Author contributions

Que Jianglong and Yu Jinxiang designed and funded this research. Rao Rongcheng and Dai Yingen performed the experiments, and analyzed data. Li Caigang and Liu Fanning analyzed part of the data. Huang Yi, Min Jialing, Wu Xiya, Shi Xinyuan and Yang Ying performed part of the experiments. Rao Rongcheng, Dai Yingen and Huang Qinghai wrote the manuscript. All authors contributed to discussing and revising.

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/s41598-025-92508-5>.

Correspondence and requests for materials should be addressed to Y.J. or Q.J.

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/>.