Ensemble modeling of black Pomfret (*Parastromateus niger*) habitat in the Taiwan Strait based on oceanographic variables

Sandipan Mondal [1, 2], **Ming An Lee** [Corresp., 1, 2, 3], **Yu-Kai Chen** [4], **Yi-Chen Wang** [1, 2]

[1] Environmental Biology & Fishery Science, National Taiwan Ocean University, Keelung, Taiwan

[2] Center of Excellence for Ocean Engineering, National Taiwan Ocean University, Keelung, Taiwan

[3] Doctoral degree program in Ocean Resource and Environmental Changes, National Taiwan Ocean University, Keelung, Taiwan

[4] Coastal and Offshore Resource Research Center, Fisheries Research Institute, Council of Agriculture, Kaohsiung, Taiwan

Corresponding Author: Ming An Lee
Email address: malee@mail.ntou.edu.tw

This study used data from Taiwan's Fisheries Agency, including fishing location, effort, captures, and time, to evaluate the geographic distribution of *Parastromateus niger* in the Taiwan Strait. Generalized linear models (GLMs) based on six oceanographic parameters outperformed other species distribution models. The locations with the greatest standardized catch-per-unit-effort (S.CPUE) of *P. niger* had a sea surface temperature (SST) of 26.5°C –29.5°C, sea surface chlorophyll (SSC) level of 0.3–0.44 mg/m^{-3}, sea surface salinity (SSS) of 33.4–34.4 psu, mixed layer depth of 10–14 m, sea surface height of 0.57–0.77 m, and eddy kinetic energy (EKE) of 0.603–0.794 $m^2/s{-}^2$. The statistical results indicate that the major factors influencing species distribution are SSS, SSC level, and EKE and that SST is only a minor factor. An ensemble habitat model was formed through the combination of four well-performing single-algorithm models with no discernible bias. The largest annual distribution of S.CPUE and nominal CPUE was located in the ranges of 117°E–119°E and 22°N–24°N.

Ensemble modeling of Black Pomfret (*Parastromateus niger*) habitat in the Taiwan Strait based on oceanographic variables

Sandipan Mondal[1,2], and Ming-An Lee[1,2,3*], Yu-Kai Chen[4] and Yi-Chen Wang[1,2]

[1]Department of Environmental Biology Fisheries Science, National Taiwan Ocean University

[2] Center of Excellence for Ocean Engineering, National Taiwan Ocean University, Keelung 20224, Taiwan

[3]Doctoral degree program in Ocean Resource and Environmental Changes, National Taiwan Ocean University, No.2, Beining Rd., Zhongzheng Dist., Keelung City 20224, Taiwan.

[4]Coastal and Offshore Resource Research Center, Fisheries Research Institute, Council of Agriculture, Keelung City 20224, Taiwan

[*] **Correspondence:** malee@mail.ntou.edu.tw

Abstract

The location, effort, number of captures, and time of fishing were all used in this study to assess the geographic distribution of Parastromateus niger in the Taiwan Strait. Other species distribution models performed worse than generalized linear models (GLMs) based on six oceanographic parameters. The sea surface temperature (SST) was between 26.5°C and 29.5°C, the sea surface chlorophyll (SSC) level was between 0.3-0.44 mg/m3, the sea surface salinity (SSS) was between 33.4°C and 34.4°C, the mixed layer depth was between 10°C and 14°C, the sea surface height was between 0.57°C and 0.77°C, and the eddy kinetic energy (EKE) was between 0.603°C According to the statistical findings, SST is merely a small effect compared to SSS, SSC level, and EKE in terms of impacting species distribution. By combining four effective single-algorithm models with no obvious bias, an ensemble habitat model was created. The ranges of 117°E-119°E and 22°N-24°N have the highest annual distributions of S.CPUE and nominal CPUE.

Keywords: boosted regression tree (BRT), classification and regression tree (CART), ensemble model, generalized additive model (GAM), generalized linear model (GLM).

1. Introduction

Species distribution models (SDMs) are the most common method of examining species habitat patterns through the use of oceanographic elements; they are also referred to as habitat models, ecological niche models, bioclimatic envelopes, and resource selection functions (Zimmerman et al., 2010, Robinson et al., 2011, Beale & Lennon, 2012, Li & Wang, 2013 and Tikhonov et al., 2020). Habitat models use mathematical representations of the current species distribution to predict the future distribution by using an algorithm. Historically, the arithmetic mean and geometric mean models (Xue et al., 2017; Li et al., 2016) based on the habitat appropriateness index have been used. However, the adoption of various machine learning models has been accelerated by technical advancement. Machine learning (ML) is a topic of study focused on

comprehending and developing "learning" techniques, or techniques that use data to improve performance on a certain set of tasks. It's thought of as a part of artificial intelligence. Without being expressly taught to do so, machine learning algorithms create a model using sample data, also referred to as training data, in order to make predictions or judgments. Machine learning algorithms are utilized in a broad variety of applications, such as in medical, email filtering, speech recognition, agriculture, and computer vision, where it is difficult or unfeasible to design traditional algorithms to do the desired tasks. Computational statistics, which focuses on making predictions with computers, is closely related to a subset of machine learning, but not all machine learning is statistical learning. The field of machine learning benefits from the tools, theory, and application domains that come from the study of mathematical optimization. Data mining is a related area of study that focuses on unsupervised learning for exploratory data analysis. Some machine learning applications employ data and neural networks in a manner that resembles how a biological brain functions. Machine learning is also known as predictive analytics when it comes to solving business problems. Different advanced models like regression models, including generalized linear models (GLMs) and generalized additive models (GAMs) as well as artificial intelligence models such as gradient boosting models (Hosain et al., 2020), artificial neural networks (Ahmad, 2019), classification and regression tree (CART) models (Youssef et al., 2016)], and random forest models (Reisinger et al., 2018) are in use present days. Ensemble models (Reisinger et al., 2021; Tabor & Koch 2021), which incorporate the forecasts of two or more habitat models (ensemble members), exhibit superior performance and robustness compared with single-algorithm models. The ensemble members may be identical or dissimilar habitat model types, and each can be trained with the same or different training sets (Georgian et al., 2019). Statistics like the mode or mean can be used to combine the ensemble members' predictions, and sophisticated techniques are employed to determine the trustworthiness of each ensemble member and the circumstances under which it can be relied upon (Abrhams et al., 2019). Many studies released in the 1990s have shown the extensive use of ensemble learning techniques such core bagging, boosting, and stacking. When a single model is insufficient, ensemble models can be used for prediction.Considering bias in the model helps lower the predictive error variance. The current study concentrated on ensemble modeling for the habitat of the black pomfret.

Black pomfret, or *Parastromateus niger* (Bloch, 1795), are small pelagic fish that live in the inshore waters of the Indian Ocean and western Pacific Ocean off the coasts of East Africa, southern Japan, and Australia (Liu, 2008). Black pomfret are found across the East China Sea and South China Sea, are prevalent in the Taiwan Strait, and play a significant role in lighting purse seine fisheries in Taiwan and in Fujian and Guangdong Provinces (Lu & Yan, 1985; Lu et al., 1991). Black pomfret production is vital to the coastal fisheries of Taiwan. However, the *P. niger* stock status in the Taiwan Strait (TS) recently collapsed (Ju et al., 2020), as evidenced by a significant drop in the annual capture in Taiwanese waters between 2002 and 2019 (Taiwan Fisheries Agency, 2019). Possible reasons behind this substantial decline in yield might be due to a combination of high demand, unregulated fishing methods, climate change, and the overfishing of *P. niger* (eg: Tao et al. 2012, Ju et al. 2020), aimed/mentioned by the United Nations Sustainable

Development Goal (SDG) 14 (Ntona & Morgera, 2018). To achieve SDG 14, overfishing must be controlled which is related to the unsustainable fish stock whereas underexploited or partially exploited levels are viable, the target of sustainable exploration.
Understanding P.niger habitat preferences and zones in-depth is essential for maintaining a species stock that is ecologically acceptable. In order to assist keep the species stock within an ecologically acceptable range, we thus used habitat modeling to evaluate the effects of the maritime environment on the P.niger habitat (Figure 1). (SDG 14.4). To stop overfishing of P.niger, a thorough study of its biological preferences and habitat regions may be helpful (SDG 14.6).

Figure 1. Study flowchart.

2. Materials and Methods

2.1 Data Collection

2.1.1. Fisheries Data

We collected information from Taiwanese fishing vessels (mostly coastal sea fishing, with gross register tonnage ranging from 0 to less than 250 tons) about *P. niger* fisheries from January 2014 to December 2019 from Taiwan's Fisheries Agency. The spatial coverage of the monthly fisheries data was 21°N–26°N and 116°E–123°E, with a resolution of 0.1°. The data provider did not specify whether the reported weights were dry or wet. Various fishing gear was used, and gears data with maximum catch contribution were only used in this study. The collected data included the year, month, latitude and longitude, catch in kilograms, effort in hours, total catch weight in one location, type of fishing gear used, and vessel identification number. Data on fishing depth and gear-soaking time were unavailable.

2.1.2. Oceanographic Data

Seven oceanographic characteristics (Table 1) were gathered from several sources for the current study: sea surface temperature (SST), sea surface salinity (SSS), mixed layer depth (MLD), sea surface chlorophyll (SSC) level, sea surface height (SSH), meridional velocity (U), and zonal velocity (V; Table 1). we calculated eddy kinetic energy (EKE) from U and V as follows: EKE = $0.5(U^2 + V^2)$. The CMEMS eddy-resolving global ocean reanalysis product GLORYS12V1 (1/12° horizontal resolution with 50 vertical levels;
https://resources.marine.copernicus.eu/product-detail/GLOBAL_MULTIYEAR_PHY_001_030/INFORMATION)
was used to collect SST, SSS, MLD, SSH, U, and V data. Its processing level and coordinate reference system are L4 and W, respectively. In addition, we gathered SSC data using the CMEMS global ocean biogeochemical hindcast product FREEGLORYS2V4 (0.25° horizontal resolution, 75 vertical levels, daily temporal resolution;
https://resources.marine.copernicus.eu/product-detail/GLOBAL_MULTIYEAR_BGC_001_029/INFORMATION),
whose processing level and coordinate reference system are Level 4 and ETRS89 (EU-recommended frame of reference for geodata for Europe 1), respectively.

These data were originally gathered between January 2014 and December 2019 and covered the geographic range of 116°E–123°E and 21°N–26°N. These data were interpolated using MATBLAB (version 2019a) to a 0.1° spatial resolution to match the fisheries data. The SSC data were interpolated to a monthly temporal resolution using MATLAB in addition to the oceanographic and fisheries data.

Table 1. Surface oceanographic data and their descriptions.

2.2. Fisheries Data Standardization

The relative abundance of *P. niger* was as assessed as the N.CPUE from a total of 55,852 observations as follows (Dunn et al., 2000; Laurdisen et al., 2008):

N.CPUE = (catch in kilograms)/(fishing effort in hours)

The use of the popular GLM standardization technique and resulting bias-filtered N.CPUE data (Hazin et al., 2007; Hinton & Maunder, 2004; Tian et al., 2009) helped to lessen the effects of spatial data, including latitude (lat) and longitude (long); temporal data (year and month); and interaction factors (i.e., year*lat, lat*long, and year*long; Mondal et al., 2022; Mondal et al., 2021; Vayghan et al., 2020; Forrestal, 2019; Shono, 2004). The key benefits of employing a GLM for standardization include the exponential distribution of response variables and the ability to employ categorical predictors. A stepwise GLM was created using the stats package in RStudio (version 3.6.0) using the aforementioned seven components (year, month, lat, long, year*lat, lat*long, and year*long). The family and procedure employed for GLM optimization were the Gaussian family and glm.fit, respectively. The GLM constructed for standardization was as follows:

$$\text{GLM: } \log(\text{N.CPUE}) = \text{year} + \text{month} + \text{lat} + \text{long} + \text{interactions},$$

where the interaction factors are year*lat, lat*long, and year*long.

2.3. S.CPUE–Oceanographic Factor Relationship

The correlations between the S.CPUE benchmark values and the aforementioned oceanographic factors were established to discern the preferred parameter ranges. We created suitability index (SI) curves for each oceanographic parameter using summed S with smoothing spline regression. The regression used S.CPUE as the dependent variable and all selected oceanographic elements as the explanatory variables. The SI curves were then normalized as follows using S.CPUE and the oceanographic variables:

$$SI = (Y - Y_{min})/(Y_{max} - Y_{min}),$$

where Y_{max} and Y_{min} are respectively the maximum and minimum number of observations of S.CPUE or oceanographic factors; thus, SI has a range between 0 and 1, where Y is a simulated or

predicted value from Ymax to Ymin. An oceanographic factor range with a large SI value (>0.6) (Mondal et al. 2021, Vayghan et al 2020, and Lee et al. 2021) suggested a favorable range for S.CPUE.

2.4. Single-Algorithm Habitat Model Development

The current study incorporated four single-algorithm models, namely a GLM, GAM, boosted regression tree (BRT) model, and CART model. Each modeling technique was optimized according to the established protocol. We developed one model for each modeling technique in RStudio and the six oceanographic factors (SST, SSS, MLD, SSH, SSC, and EKE), which were regarded as predictor variables; S.CPUE was the response variable.

We used the Gaussian family and the generalized cross-validation of the mgcv package to construct each GAM. We employed the Gaussian family and the glm.fit technique from the stats package to create each GLM. Each BRT model was built using the Gini approach and the Gaussian family from the gbm program; optimization included the use of 100 trees, seven interactions, and 0.65 bag fractions. The rpart package was used to build each CART model using the Gaussian family and the CP technique. The CART models were optimized with a CP value of 0.1 and minimum and maximum node counts of 1 and 6, respectively.

2.5. Validation of Selected Single-Algorithm Habitat Models

The fisheries data set (n = 55,852) was split into two portions using a random splitting technique performed by the RStudio caret package at a ratio of 70 [n(70) = 39,115] to 30 [n(30) = 16,737] to validate the single-algorithm models. For each single-algorithm model, three coefficients—namely the Pearson correlation coefficient (R), root-mean-square error (RMSE), and mean absolute error (MAE)—were computed for both portions of the data set. Little variation in the R, RMSE, and MAE values for the two data sets was considered indicative of a well-performing model with low bias.

2.6. Ensemble Habitat Model Development

We created an ensemble habitat model in the RStudio BIOMOD2 package (Georgian et al., 2019; Alabia et al., 2016; Reisinger et al., 2021; Tarbor & Koch, 2021; Abrahms et al., 2019) to enhance the power to predict the *P. niger* habitat. A weighted mean ensemble model of the *P. niger* habitat was created after the performance of the single-algorithm models was assessed. If no discernible bias was detected on the basis of the R, RMSE, and MAE values for the two data sets for a single-algorithm model, the model was integrated into the ensemble model. Models exhibiting potential bias were excluded.

After the creation of the ensemble habitat model, MATLAB was used to visualize the monthly value predictions of the ensemble model along with the S.CPUE for each point in the study area.

3. Results

3.1. Standardization of Fisheries Data

Over 88% of the catches were captured by otter trawl nets, gill nets, and Taiwanese seines (Figure 2). Thus only these data were selected for the analysis. The full GLM (with all six factors) resulted in an explained deviance and adjusted R^2 of 18.135% and 0.181, respectively. The residual distribution and quantile–quantile (QQ) plots (Figure 3) of the full GLM exhibited no significant fluctuation. Thus, the full GLM approach was used for the standardization of the *P. niger* fisheries data.

Figure 2. Catches from fishing with different gear.
Figure 3. Residual distributions and QQ graphs for the final GLM with predictor variables.

3.2. S.CPUE–Oceanographic Factor Relationships

The SI curves created for the S.CPUE of *P. niger* against the six oceanographic factors are illustrated in Figure 4. When the SI value exceeded 0.6, the ideal SST, SSC level, SSS, MLD, SSH, and EKE ranges were 26.5°C–29.5°C, 0.3–0.44 mg/m^3, 33.4–34.4 PSU, 10–14 m, 0.57–0.77 m, and 0.603–0.794 m^2/s^2, respectively. The total standardized CPUE of *P. niger* reached its maximum near the SST, SSC level, SSS, MLD, SSH, and EKE of 29.5°C, 0.36 mg/m^3, 34.2 PSU, 12 m, 0.67 m, and 0.661–0.724 m^2/s^2, respectively. These results indicate the preferred oceanographic range of *P. niger* in the TS during the research period.

Figure 4. Environmental ranges of *P. niger* with SI values from 2014 to 2019.

3.3. Contributions of Single Oceanographic Factors

Table 2 presents the performance of different oceanographic factors in the single-algorithm models. SSH was observed to be the most dominant oceanographic factor in all four single-algorithm models. The second most crucial oceanographic factor was EKE, which ranked second in all single-algorithm models except the GLM. SSC ranked third in all single-algorithm models except the GLM. SST, which ranked last in all the models, was deemed the least critical parameter. SSS was ranked fifth in all models but the GLM, in which it was the least influential.

Table 2. Predictive performance of six oceanographic factors in single-algorithm GAMs, GLMs, BRT models, and CART models.

3.4. Performance and Validation of Single-Algorithm Models

Table 3 presents the predictive performance of the full (with all oceanographic factors) GAM, GLM, BRT model, and CART model. The deviance explained by the full GAM, GLM, BRT model, and CART model was 54.1%, 41.52%, 52.7%, and 49.5%, respectively. The correlation values of the full GAM, GLM, BRT model, and CART model were 0.543, 0.415, 0.525, and 0.498, respectively. Figure 5a, 5b, 5c, and 5d depict the performance of the selected GAM, GLM, BRT model, and CART model, respectively.

Table 3. Predictive performance of full (all six oceanographic factors) GAM, GLM, BRT model, and CART model.

Figure 5. Residual distributions and QQ plots for diagnostic analysis of the full (a) GAM and (b) GLM, both with predictor variables. Full (c) BRT and (d) CART model performance along with the decision trees of the final model.

Small differences in the coefficients (R, RMSE, and MAE) for the two randomly apportioned data sets (70:30) indicated no significant bias in any predictive (Table 4) models, and the predictions were mapped onto a 1° geographic grid.

Table 4. Validation of selected single-algorithm models through random splitting.

3.5. Ensemble Habitat Prediction

Because no discernible bias was detected on the basis of the R, RMSE, or MAE values for the 70% and 30% portions of the data, the produced ensemble was selected for final prediction (Table 5). Figure 6 presents the predicted CPUE (P.CPUE) and S.CPUE. A high annual S.CPUE was distributed primarily in the ranges of 119°E–121°E and 21°N–25°N, the coastal waters of Taiwan. Most S.CPUE values were >4 in these locations but <1 in the remaining study areas. P.CPUE displayed a pattern indicating expansion to 26°N. Both S.CPUE and P.CPUE were between 0.1 and 5.

Table 5. Validation of ensemble model through random splitting.

Figure 6. P.CPUE from the ensemble habitat model, along with S.CPUE.

4. Discussion

The Kuroshio Current and coastal currents near Taiwan contribute to the diversity and productivity of marine species. As a result, the prevalence of fleet-based fishing operations has grown substantially throughout Taiwan's waters over the past 40 years. The fishing gear used in this region includes purse seines, bottom and pelagic trawls, longlines, and gill and set nets (Fisheries Agency, 1949–2019). However, the trend of overfishing beginning in the 1950s caused catches to peak in 1980 and gradually decline afterward (Chen et al., 2018; Liao et al., 2019). Despite frequent acknowledgment of the problematic state of coastal and offshore fisheries in Taiwanese waters (Liu, 2013; Chen, 2006; Shao et al., 2011), few fish species have been studied. Notably, the *P. niger* stocks in the waters close to Taiwan have drastically decreased (Ju et al., 2020).

4.1 Spatial Distribution

High annual S.CPUE values were observed primarily in the ranges of 119°E–121°E and 21°N–26°N. The P.CPUE values displayed a comparable pattern, with extension to 26°N. This distribution pattern may result from various factors.

First, the Kuroshio Current and coastal currents have boosted species diversity and productivity in the waters near Taiwan (Naimullah et al., 2020a). The Kuroshio Branch Current (KBC), China

Coastal Current (CCC), and South China Sea (SCS) Current are three major currents that affect the TS, which is located in the tropical to subtropical western Pacific. These currents influence the fishing grounds and marine habitats of the East China Sea and SCS that border the TS to the north and south, respectively(Naimullah et al., 2022). The KBC provides a favorable environment for the diversified *P .niger* in the TS. The CCC offers a neritic water mass with low salinity and temperature but high nutrient content because of its connection to the rivers of the Chinese mainland (Shiah et al., 2000). Contrary to popular assumption, the KBC, which is derived from the Kuroshio Current, has high salinity and temperature and a nutrient level comparable to that of the CCC (Chung et al., 2001). These traits produce a water mass with physical characteristics distinct from those of the surrounding water. Properties such as temperature and salinity affect the distribution of *P .niger*.

Second, the KBC and CCC both contribute to upwelling. The bottom current in the TS flows upward from the continental slope, and the surface current is primarily driven by wind (Naimullah et al., 2020b). In addition, the eastern side of the TS receives occasional injections of water from the Kuroshio Current. The aforementioned upwelling forces nutrient-rich, typically chilly water to ascend to the surface. The nutrients "fertilize" the surface waters and thus support a high level of biological production. Consequently, these fertilized TS zones may serve as ideal *P. niger* fishing locations.

Third, the seafloor of the TS is intricate. The seabed topography and capes influence tidal currents, which form counterclockwise eddies (Hsu et al., 2021). High chlorophyll concentrations outside the estuary are transferred by these tidal currents to the ocean current and attract secondary producers, including fish, crustaceans, and mollusks, and draw out *P. niger* for harvest.

4.2 Habitat Modeling Approach for Sustainable Development

The detailed information provided by habitat or spatial distribution modeling may assist in the sustainable management of *P. niger*. The pervasive nature of the measurement error inherent to models of species and habitat distribution may render such models unable to contribute to spatial economic optimization for sustainable planning. However, SDMs can potentially serve as heuristic tools for addressing oceanic environmental challenges. We emphasize the contextual application of such models.

Identification of fishing grounds that are underutilized or only partially utilized can be made easier using habitat models. The predicted accuracy of single-algorithm models, however, can occasionally be impacted by data changes, leading to unduly optimistic or gloomy predictions. As a result, the current work used an ensemble modeling strategy. We merged and trained several single-algorithm models, often known as weak learners, to address the same issue. Weak learners ultimately produce ensemble models that are more accurate because, despite completing tasks poorly when working alone, they collaborate with other weak learners to become strong learners.The easy identification of fishing grounds crucially enhances fisheries revenue and reduces fishing effort, travel time, fuel consumption, and cost. However, the likelihood of such simplified identification of fishing grounds to result in overfishing highlights the relevance of the

SDGs. The adoption of SDG 14 has sparked discussion about ocean health and its importance to the future of the planet (Ntona & Morgera, 2018). In here the most important aspect is the conservation. Conservation measures can be taken in the overexploited areas and SDM can be used to identify initially the distribution zone of any particular species. Condition of these high or low catch zone can be examined through stock assessment to over or underexploited areas. The SDG targets are intended address the major problems threatening ocean resources, such as overfishing and climate change, but doing so requires emphases on the socioeconomic dimensions of ocean politics and the distinct positions of the least developed countries and small island states. The SDGs have garnered institutional acceptance since their adoption (Friess et al., 2019; Sturesson et al., 2018). Understanding the habitat of *P. niger* in the TS may facilitate the sustainable management of the species. The primary aim of SDG 14.4 is biologically sustainable fish stock levels. Habitat modeling can play a crucial role in achieving this goal by identifying the *P. niger* habitat in the TS. Additionally, SDG 14.5 focuses on conservation in coastal and marine areas. Highly exploited areas can be declared protected areas through temporary fishing prohibition to promote stock sustainability. SDG 14.6 calls for an end to overfishing subsidies. Subsidies for fishing vessels traveling to less-exploited areas should be discontinued to avoid overfishing. The sustainability of the oceans and their resources can also be promoted through the enhancement of scientific understanding, research, and the transfer of marine technology. Related policies should consider the Criteria and Guidelines of the Intergovernmental Oceanographic Commission (SDG 14.a), support small-scale fisheries (SDG 14.b), and implement and uphold international maritime law (SDG 14.c). The modeling of species distribution or habitats may constitute the initial stage in sustainability research (Figure 7).

4.3 Importance of the Study

Fisheries management organizations have developed and embraced ecosystem-based management techniques in recent years, improving public awareness of the contribution of fisheries to the provision of marine ecosystem services (Kenny et al., 2018). The vital ecosystem services provided by ocean resources are fundamental to the future of the planet. Such services provide a considerable source of wealth and revenue, and their significance will continue to grow with the rising demand for land-based natural resources (Virto, 2018). However, the ability of oceans and their resources to meet the needs of their species is threatened (Neumann et al., 2017). Scientists lack a full understanding of the environmental impacts of overfishing, other destructive fishing practices, and ocean warming, which has caused rising sea levels and the expansion of ocean acidification and hypoxic or dead zones. Some of the greatest threats stem from the introduction of invasive species, primarily due to human activity. As new, capital-intensive methods of resource extraction gain popularity and less technologically advanced nations fall behind in the competition for limited maritime resources, the seas could become an arena that exacerbates global inequality (Griggs et al., 2017). As a result, many people may be forced to change or drastically reduce their demands on ocean ecosystems. Furthermore, scientists anticipate that ocean warming and escalating climate change will significantly influence the lives of millions worldwide.

The current study identified the detailed habitat preferences and zones of *P. niger* to further the maintenance of ecologically acceptable levels of species stock (SDG 14.4). A proper understanding of habitat preferences and zones can help to prevent the overfishing of *P. niger* (SDG 14.6). We plan to conduct future research on the predicted effects of climate change on *P. niger* through habitat-based modeling and to offer recommendations for sustainability.

Figure 7. Habitat modeling approach for sustainable development with SDG interactions.

5. Conclusion

This study used a variety of oceanographic characteristics to pinpoint the geographic range of P. niger in the TS. Due to the GLM approach's superior performance to other models, we chose it for standardization. Near the SST, SSC level, SSS, MLD, SSH, and EKE of 29.5°C, 0.36 mg/m^3, 34.2 PSU, 12 m, 0.67 m, and 0.661-0.724 m^2/s^2, respectively, the P. niger S.CPUE attained its highest value. According to the statistical analysis of our ensemble model, SST is the least important component and SSH and EKE are the key factors affecting the P. niger distribution. The largest yearly P.CPUE distribution followed by the largest annual S.CPUE distribution were found in the regions of 21°N-26°N and 119°E-121°E.

Author Contributions: Conceptualization, S.M. and M.A.L.; methodology, M.A.L. and S.M.; software, S.M.; validation, M.A.L.; formal analysis, S.M.; investigation, M.A.L. and Y.C.W.; resources, M.A.L. and Y.C.W.; writing (original draft preparation), S.M.; writing (review and editing), M.A.L.; visualization, Y.C.W. All authors have read and approved of the published version of the manuscript.

Funding: This research was financed by the Council of Agriculture (111AS-6.4.2-FA-F1) and the National Science and Technology Council of Taiwan (NSTC110-2621-M-019-004).

Acknowledgments: We thank the anonymous reviewers and editors for their valuable comments and suggestions, as well as the team members of the Taiwan Fisheries Agency and National Science and Technology Council of Taiwan for their assistance in data preparation.

References

1. Tikhonov, G., Opedal, Ø.H., Abrego, N., Lehikoinen, A., de Jonge, M.M., Oksanen, J. and Ovaskainen, O., 2020. Joint species distribution modelling with the R-package Hmsc. Methods in ecology and evolution, 11(3), pp.442-447.
2. Li, X. and Wang, Y., 2013. Applying various algorithms for species distribution modelling. Integrative zoology, 8(2), pp.124-13.
3. Beale, C.M. and Lennon, J.J., 2012. Incorporating uncertainty in predictive species distribution modelling. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1586), pp.247-258.
4. Robinson, L.M., Elith, J., Hobday, A.J., Pearson, R.G., Kendall, B.E., Possingham, H.P. and Richardson, A.J., 2011. Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. Global Ecology and Biogeography, 20(6), pp.789-802.

5. Zimmermann, N.E., Edwards Jr, T.C., Graham, C.H., Pearman, P.B. and Svenning, J.C., 2010. New trends in species distribution modelling. Ecography, 33(6), pp.985-989.

6. Xue, Y., Guan, L., Tanaka, K., Li, Z., Chen, Y. and Ren, Y., 2017. Evaluating effects of rescaling and weighting data on habitat suitability modeling. Fisheries Research, 188, pp.84-94.

7. Li, G., Cao, J., Zou, X., Chen, X. and Runnebaum, J., 2016. Modeling habitat suitability index for Chilean jack mackerel (*Trachurus murphyi*) in the South East Pacific. Fisheries Research, 178, pp.47-60.

8. Azzellino, A., Panigada, S., Lanfredi, C., Zanardelli, M., Airoldi, S. and di Sciara, G.N., 2012. Predictive habitat models for managing marine areas: spatial and temporal distribution of marine mammals within the Pelagos Sanctuary (Northwestern Mediterranean sea). Ocean & coastal management, 67, pp.63-74.

9. Zhang, T., Song, L., Yuan, H., Song, B. and Ebango Ngando, N., 2021. A comparative study on habitat models for adult bigeye tuna in the Indian Ocean based on gridded tuna longline fishery data. Fisheries Oceanography, 30(5), pp.584-607.

10. Hossain, M.S., Sarker, S., Sharifuzzaman, S.M. and Chowdhury, S.R., 2020. Primary productivity connects hilsa fishery in the Bay of Bengal. Scientific reports, 10(1), pp.1-16.

11. Ahmad, H., 2019. Machine learning applications in oceanography. Aquatic Research, 2(3), pp.161-169.

12. Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S. and Al-Katheeri, M.M., 2016. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides, 13(5), pp.839-856.

13. Reisinger, R.R., Raymond, B., Hindell, M.A., Bester, M.N., Crawford, R.J., Davies, D., de Bruyn, P.N., Dilley, B.J., Kirkman, S.P., Makhado, A.B. and Ryan, P.G., 2018. Habitat modelling of tracking data from multiple marine predators identifies important areas in the Southern Indian Ocean. Diversity and Distributions, 24(4), pp.535-550.

14. Reisinger, R.R., Friedlaender, A.S., Zerbini, A.N., Palacios, D.M., Andrews-Goff, V., Dalla Rosa, L., Double, M., Findlay, K., Garrigue, C., How, J. and Jenner, C., 2021. Combining regional habitat selection models for large-scale prediction: Circumpolar habitat selection of Southern Ocean humpback whales. Remote Sensing, 13(11), p.2074.

15. Tao, Y., Mingru, C., Jianguo, D., Zhenbin, L., and Shengyun, Y. (2012). Age and growth changes and population dynamics of the black pomfret (Parastromateus niger) and the frigate tuna (Auxis thazard thazard), in the Taiwan Strait. Lat. Am. J. Aquat. Res. 40, 649–656.

16. Tabor, J.A. and Koch, J.B., 2021. Ensemble models predict invasive bee habitat suitability will expand under future climate scenarios in Hawai'i. Insects, 12(5), p.443.

17. Georgian, S.E., Anderson, O.F. and Rowden, A.A., 2019. Ensemble habitat suitability modeling of vulnerable marine ecosystem indicator taxa to inform deep-sea fisheries management in the South Pacific Ocean. Fisheries research, 211, pp.256-274.

18. Abrahms, B., Welch, H., Brodie, S., Jacox, M.G., Becker, E.A., Bograd, S.J., Irvine, L.M., Palacios, D.M., Mate, B.R. and Hazen, E.L., 2019. Dynamic ensemble models to predict distributions and anthropogenic risk exposure for highly mobile species. Diversity and Distributions, 25(8), pp.1182-1193.

19. Reisinger, R.R., Raymond, B., Hindell, M.A., Bester, M.N., Crawford, R.J., Davies, D., de Bruyn, P.N., Dilley, B.J., Kirkman, S.P., Makhado, A.B. and Ryan, P.G., 2018. Habitat modelling of tracking data from multiple marine predators identifies important areas in the Southern Indian Ocean. Diversity and Distributions, 24(4), pp.535-550.

20. Rowden, A.A., Anderson, O.F., Georgian, S.E., Bowden, D.A., Clark, M.R., Pallentin, A. and Miller, A., 2017. High-resolution habitat suitability models for the conservation and management of vulnerable marine ecosystems on the Louisville Seamount Chain, South Pacific Ocean. Frontiers in Marine Science, 4, p.335.

21. Alabia, I.D., Saitoh, S.I., Igarashi, H., Ishikawa, Y., Usui, N., Kamachi, M., Awaji, T. and Seito, M., 2016. Ensemble squid habitat model using three-dimensional ocean data. ICES Journal of Marine Science, 73(7), pp.1863-1874.

22. Liu, R.Y. 2008. Checklist of marine biota of China Seas. Science Press, Beijing, 1267 pp.

23. Lu, Z. & Y. Youming. 1985. Age and growth of Parastromateus niger in Western Taiwan Strait. J. Fujian Fish., 3: 7-13.

24. Lu, Z., D. Quanshui & Y. Youming. 1991. Growth and mortality of Auxis thazard in the Taiwan Strait and its adjacent sea. J. Fish. China, 15(3): 228-235.

25. Taiwan Fisheries Agency, 2019.

26. Ntona, M. and Morgera, E., 2018. Connecting SDG 14 with the other Sustainable Development Goals through marine spatial planning. Marine Policy, 93, pp.214-222.

27. Cormier, R. and Elliott, M., 2017. SMART marine goals, targets and management–is SDG 14 operational or aspirational, is 'Life Below Water'sinking or swimming? Marine pollution bulletin, 123(1-2), pp.28-33.

28. Mugagga, F. and Nabaasa, B.B., 2016. The centrality of water resources to the realization of Sustainable Development Goals (SDG). A review of potentials and constraints on the African continent. International Soil and Water Conservation Research, 4(3), pp.215-223.

29. Ju, P., Tian, Y., Chen, M., Yang, S., Liu, Y., Xing, Q. and Sun, P., 2020. Evaluating stock status of 16 commercial fish species in the coastal and offshore waters of Taiwan using the CMSY and BSM methods. Frontiers in Marine Science, 7, p.618.

30. Dunn, A., Harley, S.J., Doonan, I.J. and Bull, B., 2000. Calculation and interpretation of catch-per-unit effort (CPUE) indices. New Zealand fisheries assessment report, 1, p.44.

31. Lauridsen, T.L., Landkildehus, F., Jeppesen, E., Jørgensen, T.B. and Søndergaard, M., 2008. A comparison of methods for calculating Catch Per Unit Effort (CPUE) of gill net catches in lakes. Fisheries Research, 93(1-2), pp.204-211.

32. Hazin, H.G., Hazin, F., Travassos, P., Carvalho, F.C. and Erzini, K., 2007. Standardization of swordfish CPUE series caught by Brazilian longliners in the Atlantic Ocean, by GLM, using the targeting strategy inferred by cluster analysis. Collect. Vol. Sci. Pap. ICCAT, 60(6), pp.2039-2047.

33. Hinton, M.G. and Maunder, M.N., 2004. Methods for standardizing CPUE and how to select among them. Col. Vol. Sci. Pap. ICCAT, 56(1), pp.169-177.

34. Tian, S., Chen, X., Chen, Y., Xu, L. and Dai, X., 2009. Standardizing CPUE of Ommastrephes bartramii for Chinese squid-jigging fishery in Northwest Pacific Ocean. Chinese Journal of Oceanology and Limnology, 27(4), pp.729-739.

35. Mondal, S., Vayghan, A.H., Lee, M.A., Wang, Y.C. and Semedi, B., 2021. Habitat suitability modeling for the feeding ground of immature albacore in the southern Indian Ocean using satellite-derived sea surface temperature and chlorophyll data. Remote Sensing, 13(14), p.2669.

36. Vayghan, A.H., Lee, M.A., Weng, J.S., Mondal, S., Lin, C.T. and Wang, Y.C., 2020. Multisatellite-based feeding habitat suitability modeling of albacore tuna in the southern Atlantic ocean. Remote Sensing, 12(16), p.2515.

37. Mondal, S., Wang, Y.C., Lee, M.A., Weng, J.S. and Mondal, B.K., 2022. Ensemble Three-Dimensional Habitat Modeling of Indian Ocean Immature Albacore Tuna (Thunnus alalunga) Using Remote Sensing Data. *Remote Sensing*, *14*(20), p.5278..

38. Forrestal, F.C., Schirripa, M., Goodyear, C.P., Arrizabalaga, H., Babcock, E.A., Coelho, R., Ingram, W., Lauretta, M., Ortiz, M., Sharma, R. and Walter, J., 2019. Testing robustness of CPUE standardization and inclusion of environmental variables with simulated longline catch datasets. Fisheries research, 210, pp.1-

13.

39. Shono, H., 2004. A review of some statistical approaches used for CPUE standardization. Bulletin of the Japanese Society of Fisheries Oceanography (Japan).

40. Georgian, S.E., Anderson, O.F. and Rowden, A.A., 2019. Ensemble habitat suitability modeling of vulnerable marine ecosystem indicator taxa to inform deep-sea fisheries management in the South Pacific Ocean. Fisheries research, 211, pp.256-274.

41. Alabia, I.D., Saitoh, S.I., Igarashi, H., Ishikawa, Y., Usui, N., Kamachi, M., Awaji, T. and Seito, M., 2016. Ensemble squid habitat model using three-dimensional ocean data. ICES Journal of Marine Science, 73(7), pp.1863-1874.

42. Reisinger, R.R., Friedlaender, A.S., Zerbini, A.N., Palacios, D.M., Andrews-Goff, V., Dalla Rosa, L., Double, M., Findlay, K., Garrigue, C., How, J. and Jenner, C., 2021. Combining regional habitat selection models for large-scale prediction: Circumpolar habitat selection of Southern Ocean humpback whales. Remote Sensing, 13(11), p.2074.

43. Tabor, J.A. and Koch, J.B., 2021. Ensemble models predict invasive bee habitat suitability will expand under future climate scenarios in Hawai'i. Insects, 12(5), p.443.

44. Abrahms, B., Welch, H., Brodie, S., Jacox, M.G., Becker, E.A., Bograd, S.J., Irvine, L.M., Palacios, D.M., Mate, B.R. and Hazen, E.L., 2019. Dynamic ensemble models to predict distributions and anthropogenic risk exposure for highly mobile species. Diversity and Distributions, 25(8), pp.1182-1193.

45. Chen, J.L., Lin, Y.S. and Chuang, C.T., 2018. Improving the management of Taiwanese fishery resource conservation zones based on public perceptions and willingness to pay for ecosystem services. Journal of Coastal Conservation, 22(2), pp.385-398.

46. Liao, C.P., Huang, H.W. and Lu, H.J., 2019. Fishermen's perceptions of coastal fisheries management regulations: Key factors to rebuilding coastal fishery resources in Taiwan. Ocean & Coastal Management, 172, pp.1-13.

47. Liu, W.H., 2013. Managing the offshore and coastal fisheries in Taiwan to achieve sustainable development using policy indicators. Marine Policy, 39, pp.162-171.

48. Chen, J.L., 2006. A study on the planning and management of fishery resource conservation areas in Taiwan. National Taiwan Ocean University, Keelung (in Chinese).

49. Shao, K.T., Soong, K.Y., Lin, C.W., Wu, S.P., Chan, T.Y. and Chang, J.S., 2011. Investigation and planning of fishery resources conservation zones and rare species. Fishery Agency, Taipei (in Chinese).

50. Kenny, A.J., Campbell, N., Koen-Alonso, M., Pepin, P. and Diz, D., 2018. Delivering sustainable fisheries through adoption of a risk-based framework as part of an ecosystem approach to fisheries management. Marine Policy, 93, pp.232-240.

51. Ntona, M. and Morgera, E., 2018. Connecting SDG 14 with the other Sustainable Development Goals through marine spatial planning. Marine Policy, 93, pp.214-222.

52. Friess, D.A., Aung, T.T., Huxham, M., Lovelock, C., Mukherjee, N. and Sasmito, S., 2019. SDG 14: life below water–impacts on mangroves. Sustainable Development Goals, 445, pp.445-481.

53. Sturesson, A., Weitz, N. and Persson, Ã., 2018. SDG 14: Life Below Water. A Review of Research Needs. Technical annex to the Formas report ForskningfÃ¶ r Agenda, 2030.

54. Virto, L.R., 2018. A preliminary assessment of the indicators for Sustainable Development Goal (SDG) 14 "Conserve and sustainably use the oceans, seas and marine resources for sustainable development". Marine Policy, 98, pp.47-57.

55. Neumann, B., Ott, K. and Kenchington, R., 2017. Strong sustainability in coastal areas: a conceptual interpretation of SDG 14. Sustainability science, 12(6), pp.1019-1035.

541 56. Griggs, D.J., Nilsson, M., Stevance, A. and McCollum, D., 2017. A guide to SDG interactions: from science
542 to implementation. International Council for Science, Paris.

543 57. Naimullah, M., Lan, K.W., Liao, C.H., Hsiao, P.Y., Liang, Y.R. and Chiu, T.C., 2020. Association of
544 environmental factors in the taiwan strait with distributions and habitat characteristics of three swimming
545 crabs. Remote Sensing, 12(14), p.2231.

546 58. Naimullah, M., Lee, W.Y., Wu, Y.L., Chen, Y.K., Huang, Y.C., Liao, C.H. and Lan, K.W., 2022. Effect of
547 soaking time on targets and bycatch species catch rates in fish and crab trap fishery in the southern East China
548 Sea. Fisheries Research, 250, p.106258.

549 59. Shiah, F.K., Chung, S.W., Kao, S.J., Gong, G.C. and Liu, K.K., 2000. Biological and hydrographical
550 responses to tropical cyclones (typhoons) in the continental shelf of the Taiwan Strait. Continental Shelf
551 Research, 20(15), pp.2029-2044.

552 60. Chung, S.W., Jan, S. and Liu, K.K., 2001. Nutrient fluxes through the Taiwan Strait in spring and summer
553 1999. Journal of Oceanography, 57(1), pp.47-53.

554 61. Naimullah, M., Lan, K.W., Liang, Y.R., Hsiao, P.Y., Chiu, T.C. and Liao, C.H., 2020. Distribution and
555 habitat characteristics of three important commercial swimming crab (Crustacea: Decapoda: Portunidae)
556 related with the environmental factors in Taiwan Strait.

Table 1(on next page)

Surface oceanographic data sources and their descriptions

1 **Table 1.** Surface oceanographic data sources and their descriptions.

Environmental data	Abbreviation	Unit	Spatial resolution	Temporal resolution
Sea surface temperature	SST	°C	0.083° × 0.083°	Monthly
Sea surface salinity	SSS	psu		
Mixed layer depth	MLD	m		
Sea surface height above geoid	SSH			
Meridional velocity	U	ms^{-1}		
Zonal velocity	V			
Sea surface chlorophyll	SSC	mgm^{-3}	0.25° × 0.25°	Daily

2

Table 2(on next page)

The performance of different oceanographic factors in approaches based on GAMs, GLMs, BRTs, and CARTs

Table 2. The performance of different oceanographic factors in approaches based on GAMs, GLMs, BRTs, and CARTs.

Single-Algorithm	SST		SSC		SSS		MLD		SSH		EKE	
	Dev.	*Rank*	*Dev.*	*Rank*	*Dev.*	*Rank*	*Dev.*	*Rank*	*Dev.*	*Rank*	*Dev.*	*Rank*
GAM	7.68	**6**	19.4	3	13.7	5	13.8	4	33.5	**1**	21.6	2
GLM	1.41	5	12.27	2	0.01	**6**	11.76	3	31.34	**1**	10.74	4
BRT	9.3	**6**	20.9	3	14.1	5	16.1	4	34.6	**1**	22.1	2
RF	8.9	**6**	18.5	3	10.9	5	15.1	4	33.2	**1**	20.6	2

Table 3(on next page)

The predictive performance of full models (with all oceanographic factors) of GAM, GLM, BRT, and CART

Table 3. The predictive performance of full models (with all oceanographic factors) of GAM, GLM, BRT, and CART.

Single-Algorithm	R^2	Dev. Exp. (%)
GAM	0.543	54.1
GLM	0.415	41.52
BRT	0.525	52.7
CART	0.498	49.5

Table 4(on next page)

Validation of selected single-algorithm models through random splitting

Table 4. Validation of selected single-algorithm models through random splitting.

Single algorithm	*70% (n = 39,115)*			*30% (n = 16,737)*		
	R^2	**RMSE**	**MAE**	**R^2**	**RMSE**	**MAE**
GAM	0.479	1.452	1.344	0.482	1.448	1.343
GLM	0.361	1.454	1.331	0.345	1.455	1.332
BRT	0.528	0.503	0.373	0.527	0.502	0.371
CART	0.496	0.516	0.387	0.521	0.504	0.372

Table 5(on next page)

Ensemble habitat model’s validation through random splitting

Table 5. Ensemble habitat model's validation through random splitting.

70% (n = 39,115)			*30% (n = 16,737)*		
R^2	**RMSE**	**MAE**	**R^2**	**RMSE**	**MAE**
0.624	1.311	1.475	0.621	1.333	1.483

Figure 1

Study flowchart

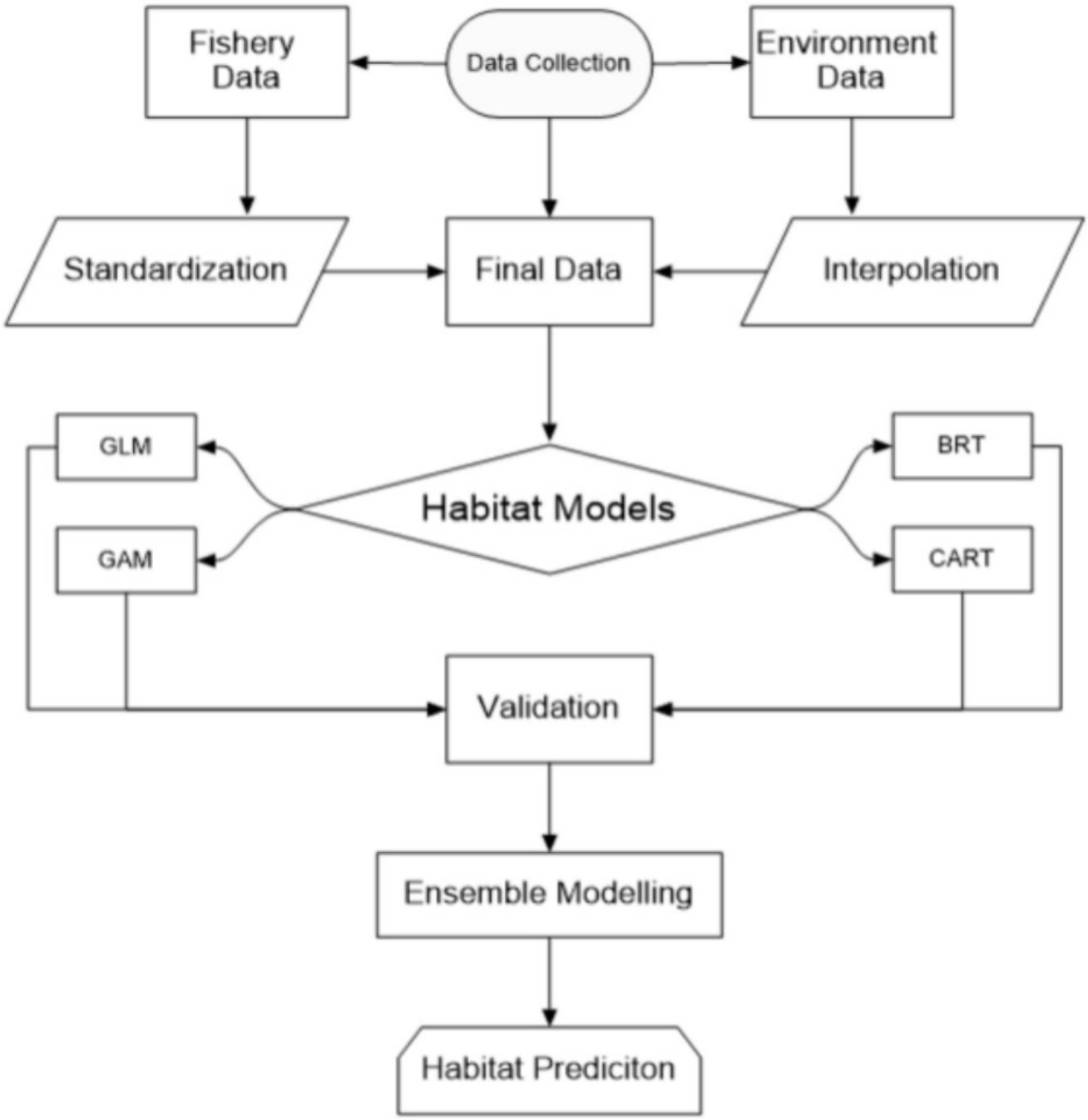


Figure 2

Catches from fishing with different gear

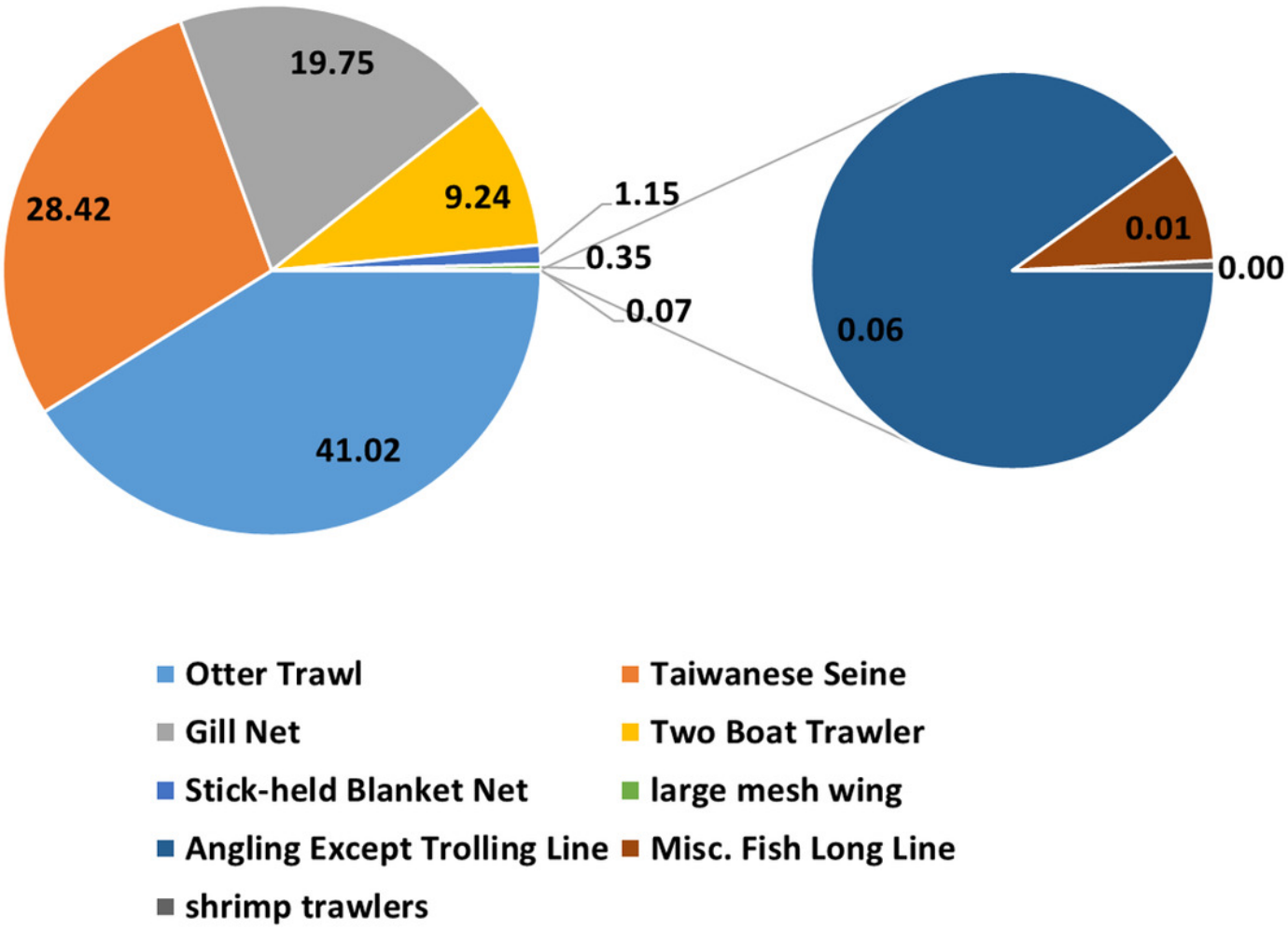
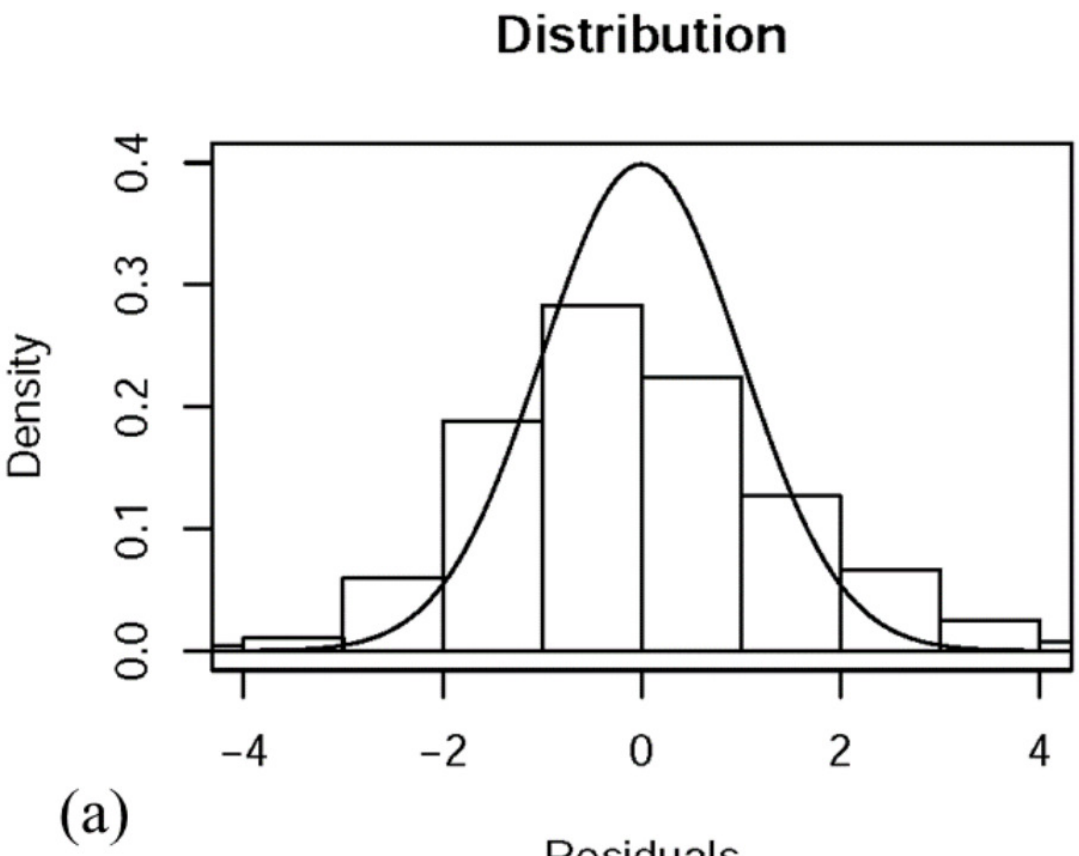
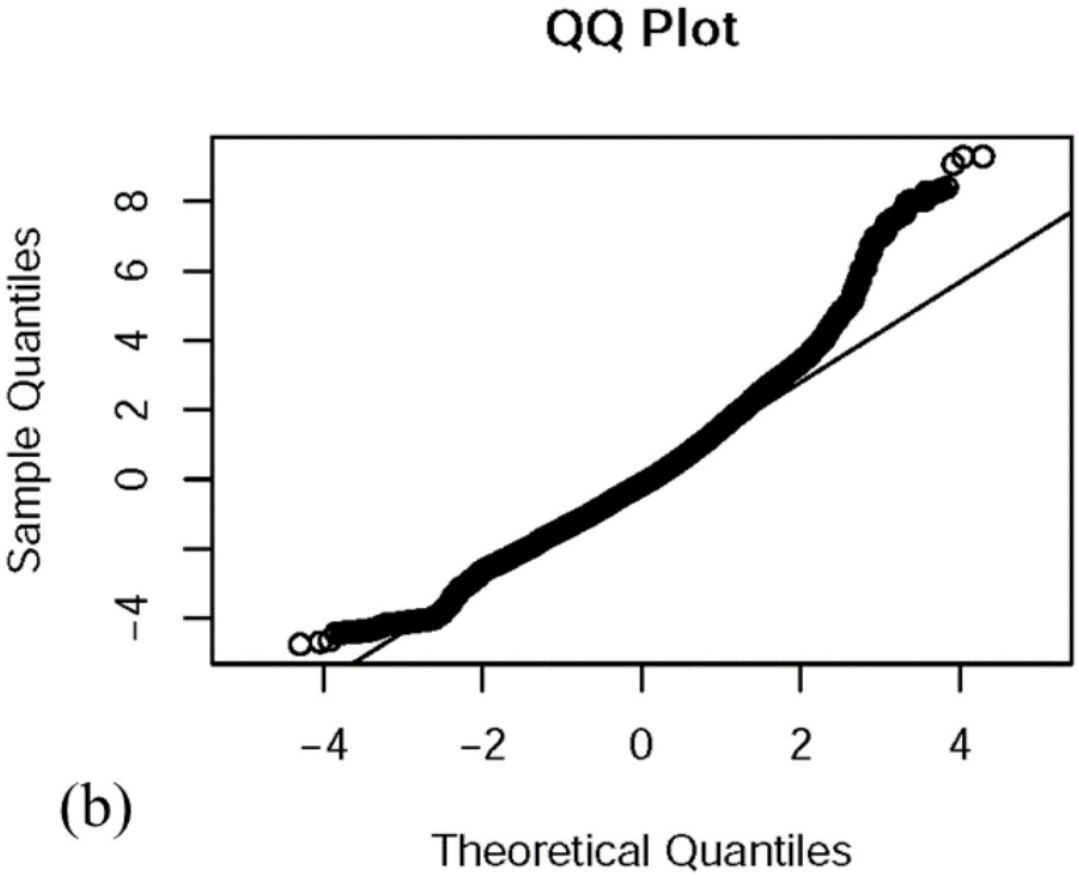


Figure 3

Residual distributions and QQ graphs for the final GLM with predictor variables


Distribution
Density
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
Residuals

(a)


QQ Plot
Sample Quantiles
−4
0
2
4
6
8
−4
−2
0
2
4
Theoretical Quantiles

(b)

Figure 4

Environmental ranges of *P. niger* with SI values from 2014 to 2019

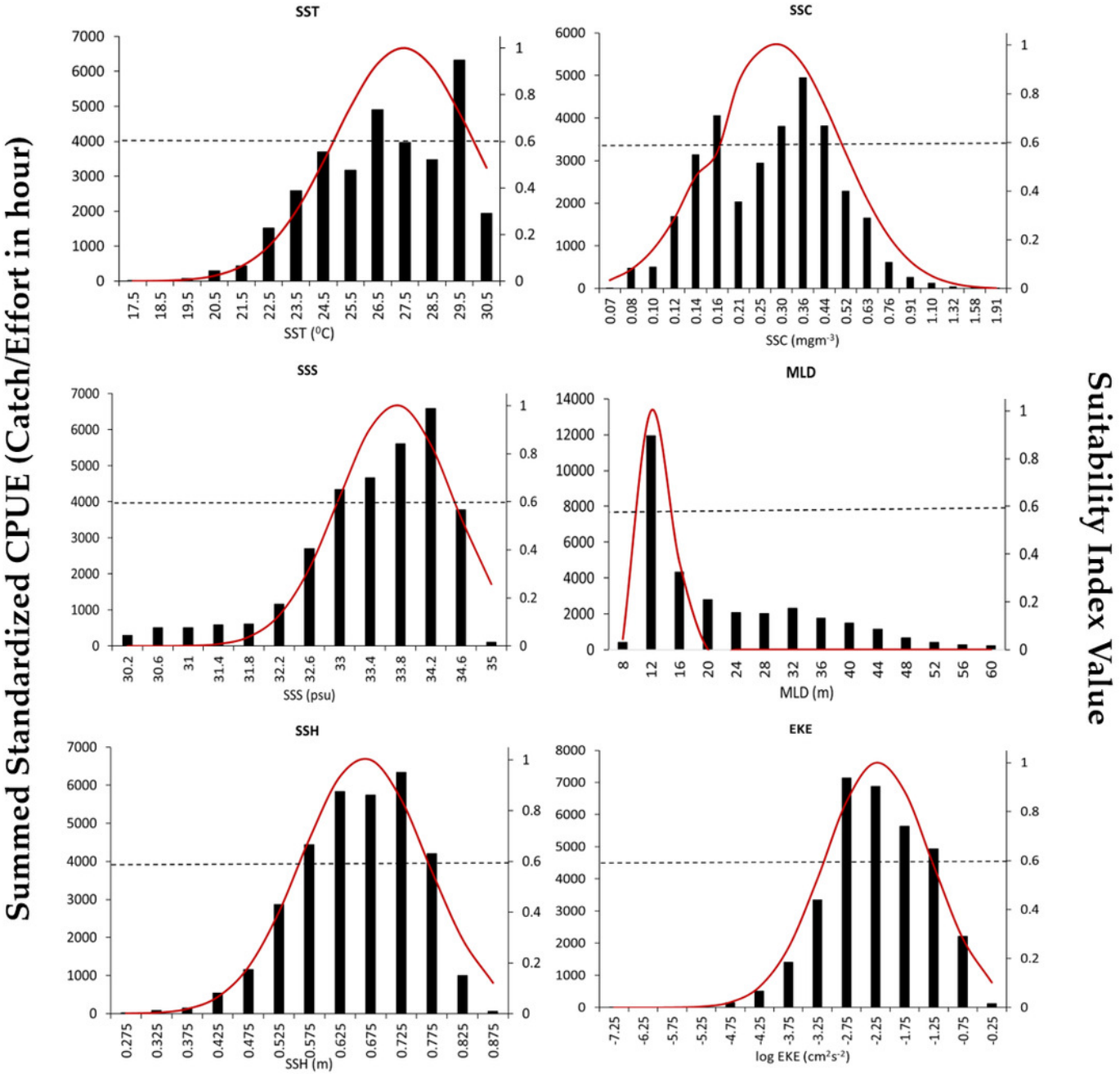
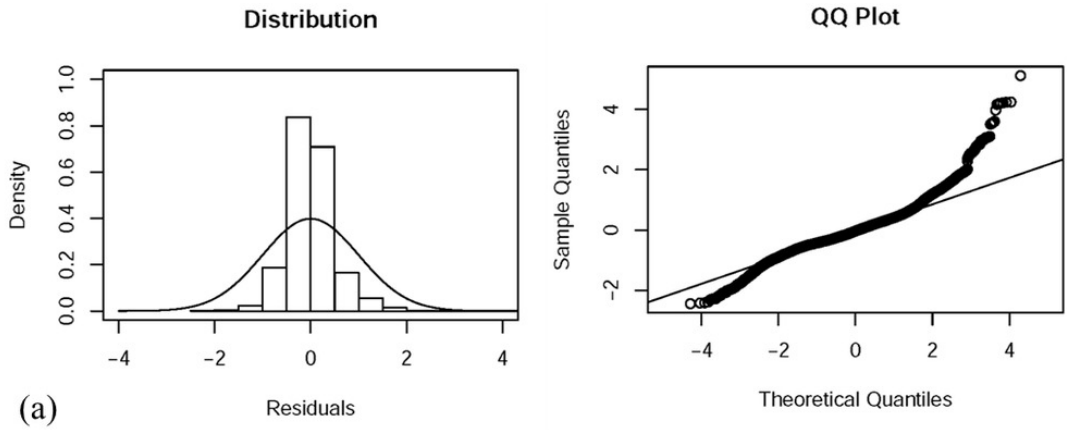
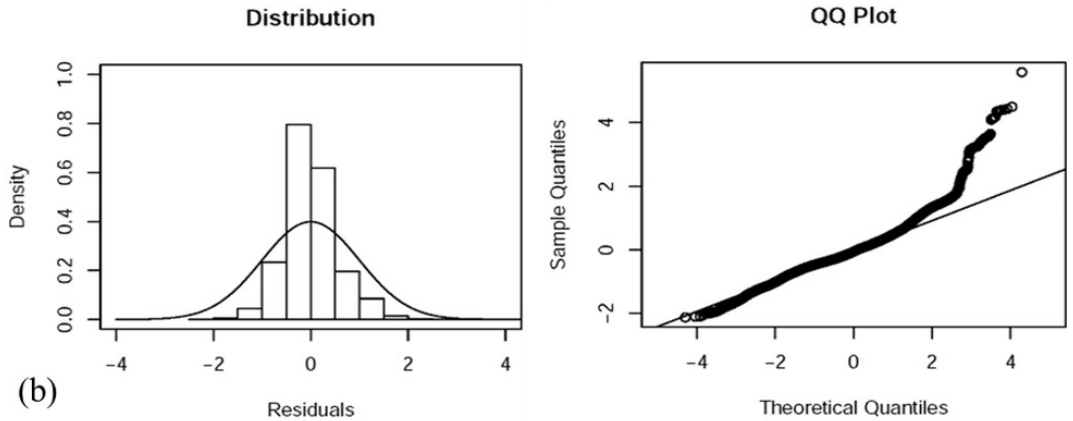


Figure 5

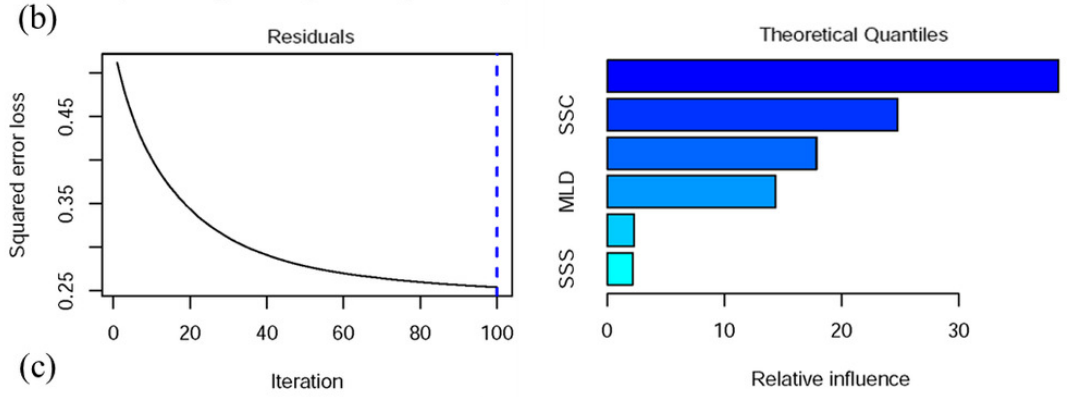
Residual distributions and QQ plots for diagnostic analysis of the full (a) GAM and (b) GLM, both with predictor variables. Full (c) BRT and (d) CART model performance along with the decision trees of the final model.


Distribution
Density
Residuals
QQ Plot
Sample Quantiles
Theoretical Quantiles

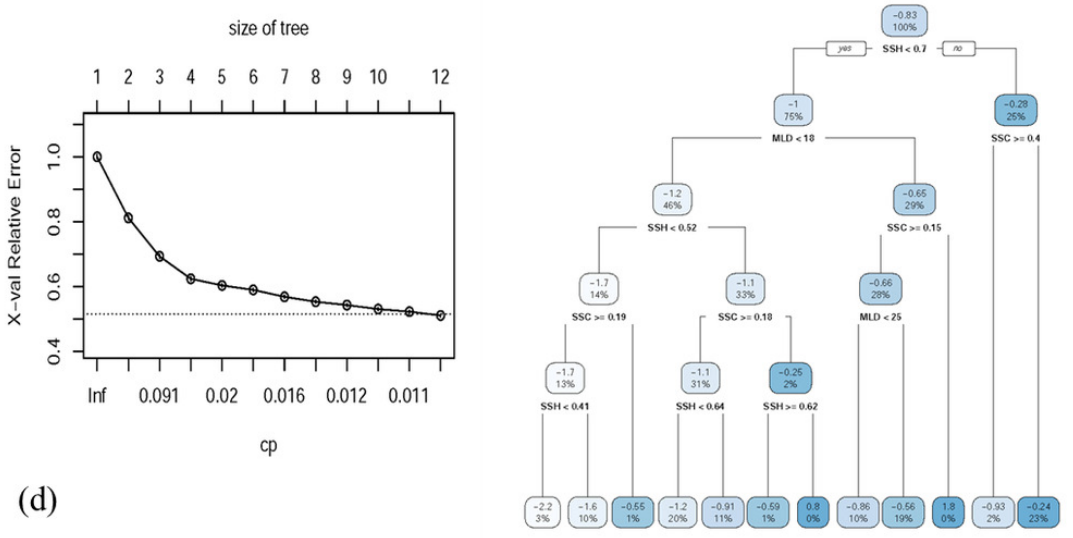
(a)


Distribution
Density
QQ Plot
Sample Quantiles

(b)


Residuals
Squared error loss
Iteration
Theoretical Quantiles
SSC
MLD
SSS
Relative influence

(c)


size of tree
X-val Relative Error
cp
SSH < 0.7
MLD < 18
SSC >= 0.4
SSH < 0.52
SSC >= 0.15
SSC >= 0.19
SSC >= 0.18
MLD < 25
SSH < 0.41
SSH < 0.64
SSH >= 0.62

(d)

Figure 6

P.CPUE from the ensemble habitat model, along with S.CPUE.

(a) S.CPUE, (b) P.CPUE

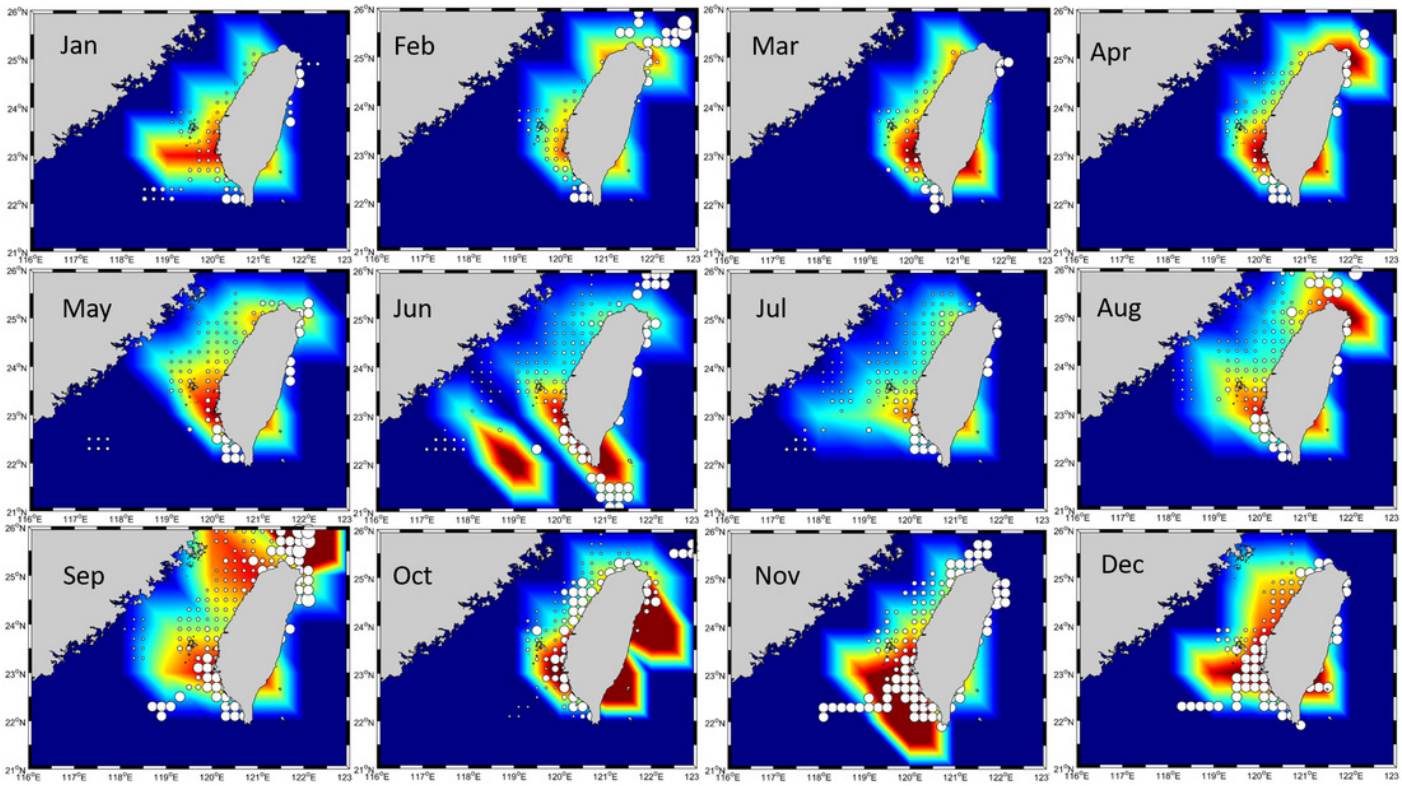


Figure 7

Habitat modeling approach for sustainable development with SDG interactions