Acta Oceanol. Sin., 2022, Vol. 41, No. 6, P. 22–30
https://doi.org/10.1007/s13131-021-1912-1
http://www.aosocean.com
E-mail: ocean2@hyxb.org.cn

Estimating seasonal habitat suitability for migratory species in the Bohai Sea and Yellow Sea: A case study of Tanaka's snailfish (*Liparis tanakae*)

Yunlong Chen[1, 2], Xiujuan Shan[1, 2*], Dingyong Zeng[3], Harry Gorfine[4], Yinfeng Xu[3], Qiang Wu[1, 2], Tao Yang[1, 2], Xianshi Jin[1, 2]

[1] Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture and Rural Affairs, Qingdao 266071, China

[2] Laboratory for Marine Fisheries Science and Food Production Processes, Pilot National Laboratory for Marine Science and Technology (Qingdao), Qingdao 266237, China

[3] State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China

[4] School of Biosciences, The University of Melbourne, Parkville, Victoria 3010, Australia

Received 14 June 2021; accepted 3 August 2021

Abstract

Acquiring a comprehensive and accurate understanding of habitat preference is essential for species conservation and fishery management, especially for mobile species that migrate seasonally. Presence and absence data from field surveys are recommended when available due to their high reliability. Using field survey data, we investigated seasonal habitat suitability requirements for Tanaka's snailfish (*Liparis tanakae*) in the Bohai Sea and Yellow Sea (BSYS) via a machine-learning method, random forests (RFs). Five environmental and biologically relevant variables (bottom temperature, bottom salinity, current velocity, depth and distance to shore) were used to build the ecological niches between the presence/absence data and suitable habitat. In addition, the degree to which false absence data might impact model performance was evaluated. Our results indicated that RFs provided accurate predictions, with seasonal habitat suitability maps of *L. tanakae* differing substantially. Bottom temperature and salinity were identified as important factors influencing the distribution of *L. tanakae*. False absence data were found to have negative effects on model performance and the decrease in evaluation metrics was usually significant (P<0.05) after 30% or more errors were added to the absence data. Through identifying highly suitable areas within its geographic range, our study provides a baseline for *L. tanakae* that can be further applied in ecosystem modelling and fishery management in the BSYS.

Key words: species distribution model, fishery-independent survey, model performance, range prediction, cold-temperate species

Citation: Chen Yunlong, Shan Xiujuan, Zeng Dingyong, Gorfine Harry, Xu Yinfeng, Wu Qiang, Yang Tao, Jin Xianshi. 2022. Estimating seasonal habitat suitability for migratory species in the Bohai Sea and Yellow Sea: A case study of Tanaka's snailfish (*Liparis tanakae*). Acta Oceanologica Sinica, 41(6): 22–30, doi: 10.1007/s13131-021-1912-1

1 Introduction

Acquiring a comprehensive and accurate understanding of habitat preference is essential for species conservation and fishery management (Beck et al., 2020; Tanaka et al., 2020). As a basic characteristic, patterning in spatial distribution aids interpretation of intra/inter-specific population dynamics. Through establishing relationships between spatial distribution and environmental variables, a habitat suitability index has been widely applied to assess habitat quality and to predict potential effects caused by human activities and climate change on commercially and ecologically important species (Lauria et al., 2015; Wisz et al., 2015; Phillips et al., 2017). A wide range of modelling techniques has been developed to deal with this issue including regression, classification, and machine-learning methods (Barbet-Massin et al., 2012). These modelling techniques statistically link the species observation data with environmental variables at different temporal and spatial scales. Among them, machine-learning methods such as the random forests (RFs) and generalized boosting models (GBMs) are popular due to their ease of implementation and superior model performance (Elith et al., 2008; Marx and Quillfeldt, 2018; Sarquis et al., 2018). Most species distribution models (SDMs) require both presence and absence data to map species distributions, as there are only a few presence-only models such as the bioclimatic envelope model (Hao et al., 2019). Although large-scale field surveys can be expensive, they provide a reliable and practical way to obtain accurate data

Foundation item: The National Natural Science Foundation of China under contract No. 42176151; the Youth Talent Program Supported by Laboratory for Marine Fisheries Science and Food Production Processes, Pilot National Laboratory for Marine Science and Technology (Qingdao) under contract No. 2018-MFS-T05; the Central Public-Interest Scientific Institution Basal Research Fund, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences under contract Nos 20603022019010 and 20603022022022.
*Corresponding author, E-mail: shanxj@ysfri.ac.cn

about species distributions, inclusive of species absence data. When using the statistical algorithms, absence data (true absence or pseudo-absence) facilitates identification of favorable conditions as opposed to relying solely on presence data (Brotons et al., 2004). Absence data obtained from field surveys may provide more information than just lack of habitat suitability. Although it is common that species absence can be caused by unfavorable environmental conditions, it can also arise from localized species extinctions, fragmentation of suitable patches of habitat, and reduced catchability of the target species (Gibson et al., 2007; Lobo et al., 2010; Chen et al., 2021). Nevertheless, false absence data can introduce substantial bias to SDMs, leading to poor performance when fitting models (Gibson et al., 2007). Therefore, it is necessary to evaluate the degree to which false absence data can impact the accuracy of habitat suitability assessments.

The Bohai Sea and Yellow Sea (BSYS) form a traditional fishing zone in China, serving as a vital spawning, nursing, feeding and over-wintering ground for numerous commercial marine species, e.g., small yellow croaker (*Larimichthys polyactis*), Japanese Spanish mackerel (*Scomberomorus niphonius*) and Pacific cod (*Gadus macrocephalus*) (Jin and Tang, 1996; Chen et al., 2018). Water circulation in the BSYS features longshore currents, the Yellow Sea Cold Water Mass (YSCWM) and the Yellow Sea Warm Current (YSWC), among which YSWC characterized by higher temperature and salinity is the only source of flow into the BSYS from the open sea (Wang and Liu, 2009; Zhong et al., 2018). This special hydrological condition makes faunal characteristics of fishery resources in this region unique. Most species live within the BSYS for all their life history stages, with regular seasonal movements forming relatively independent and diverse geographical populations.

Rapid development of marine ecosystem modelling in recent years has created an urgent and challenging demand for information about temporal and spatial distributions of commercially and ecologically important species at high resolution (Fu et al., 2017). Importantly, most species in the BSYS belong to migratory species and their distribution characteristics vary seasonally. Poor understanding of the temporal and spatial ecology of these species may cause sizeable discrepancies among results from different approaches to ecosystem modelling of the BSYS. Therefore, it is necessary to precisely delineate spatial distribution patterns and their temporal variations to identify the environmental factors driving the population dynamics.

Tanaka's snailfish (*Liparis tanakae*) is a cold-temperate species widely distributed in the Northwest Pacific including the Yellow Sea, Bohai Sea, East China Sea, Sea of Japan and Sea of Okhotsk (Jin et al., 2003; Tomiyama et al., 2013a). Characterized by seasonal migration, Tanaka's snailfish prefers to inhabit muddy bottom substrates at water depths of 50–90 m (Zhou et al., 2012). Over several decades, *L. tanakae* has become one of the top predators in the Yellow Sea ecosystem due to the dual impacts of environmental changes and fishing activities (Chen et al., 2018). Currently, studies of *L. tanakae* mainly concentrate on relative stock density, biological and reproductive characteristics, feeding ecology and helminth parasites (Jin et al., 2010; Guo et al., 2014; Park et al., 2017; Chen et al., 2018), whereas the distribution patterns and potential relationship with environment for this species are still unclear. Despite having no commercial value (Chernova et al., 2004), it is necessary to study the seasonal migration pattern of *L. tanakae* and the key ecological traits which may affect its distribution, considering its high ecological value and the vital role it plays in the BSYS ecosystem.

In this present work, we aimed at (1) assessing seasonal habitat suitability for *L. tanakae* in terms of the main environmental variables in the BSYS; and (2) evaluating the potential impact that sampling bias might have on the performance of SDMs using presence/absence data from fishery-independent surveys. Our study provides insights illustrating the relationship between species biogeography and environment of *L. tanakae* and assists in providing a scientific basis and guidance for further ecosystem studies.

2 Materials and methods

2.1 *Study area and data collection*

The BSYS are marginal seas located in the western Pacific Ocean with a total area of 460 000 km^2 (Fig. 1). Presence/absence data of *L. tanakae* were collected by seasonal fishery-independent surveys conducted by the Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences using the R/V *Beidou* in 2016. Parameters of the fishing gear were as follows: a net of circumference 836 mesh×12 cm, and a 10 cm mesh-size cod-end with a 2.4 cm mesh-size liner thereafter. The duration of each trawl shot varied between 0.5 h and 1 h at an average hauling speed of 3 kn. All data were standardized to 1 h trawl duration for further analyses. The numbers of survey sites were 48 (January), 74 (June), 98 (August) and 96 (October), respectively. Although the number of survey sites varied between seasons, relatively high sampling coverage was achieved throughout the entire geographic range during each survey.

Considering both biological relevance and data availability, we chose five environmental parameters (bottom temperature, bottom salinity, bottom water velocity in longitudinal and latitudinal direction, water depth, and distance to shore) as the predictive variables to infer the habitat distribution of *L. tanakae*. Distance to shore was extracted from the Global Marine Environment Datasets (http://gmed.auckland.ac.nz) (Basher et al., 2018). Depth, monthly mean bottom temperature, bottom salinity and bottom velocity in 2016 were obtained from the Regional Ocean Modeling Systems, physical model provided by the Second Institute of Oceanography, Ministry of Natural Re-

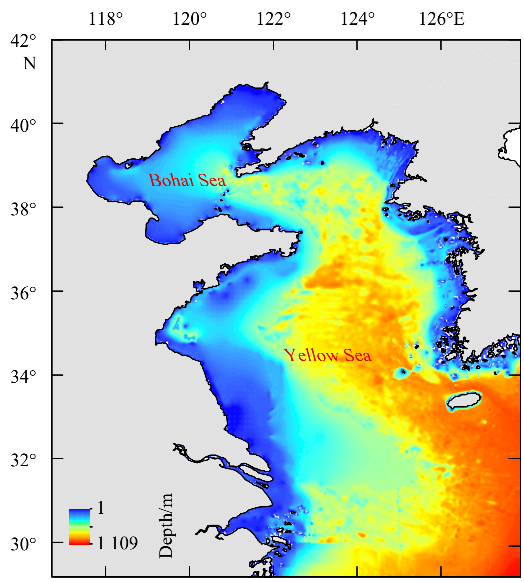

Fig. 1. Study area.

sources. These data have been validated with direct observations and other published data from reanalysis, such as the NOAA optimum interpolation sea surface temperature and the Hybrid Coordinate Ocean Model data, which are commonly accepted to be valid and reliable (Figs S1–S3). For modelling purposes, all environmental data were cropped to the study area (32°–41°N, 117°–127°E) at a spatial resolution of 5′. Among the environmental variables tested, depth and distance to shore were treated as static variables. To avoid possible multicollinearity which could lead to biased model estimation, Pearson's correlation coefficients were compared among environmental factors. Variables with an absolute correlation coefficient value higher than 0.7 were discarded before further analysis (Schickele et al., 2020). Depth was omitted in all four seasons due to its strong correlation with bottom temperature and bottom salinity. Bottom salinity in January was omitted due to its high correlation with bottom temperature (Fig. S4).

2.2 *Model performance and evaluation*

Seasonal habitat suitability for *L. tanakae* was estimated using the RFs method and applied to the presence/absence data mentioned in Section 2.1. The RFs modelling technique was chosen due to its robust performance and increased application in recent years (Melnychuk et al., 2017; Pons et al., 2018; Rubio et al., 2020). It is a machine-learning approach based on the classification and regression tree algorithm. It combines bagging and random selection of variables together, builds a "forest" consisting of a selected number of classification or regression trees and uses the voting method to obtain the final result (Breiman, 2001). For each tree, a set of variables are randomly selected from the original variables at a given node where the data are split. Further details of the RFs method are more fully described in Breiman (2001) and Cutler et al. (2007).

Considering the reliability of presence data from actual observation, we focused more on addressing uncertainty in the absence data. Due to a lack of more detailed information about the survey absence data, here we assume that the absence data are more likely to consist of both "actual" and "false" absence data. For each sampling season, experimental groups with false absence data were created by changing proportions of absence to presence through adding different levels of errors randomly (0, 10%, 20%, 30%, 40% and 50%). In this way, the effect of uncertainty in absence data could be measured by evaluating model predictions with the control group. Mean difference and statistical significance (a=0.05) of evaluation metrics for different groups (a control group with no error added and experimental groups with different levels of error added) were tested by either one-way ANOVA or Kruskal-Wallis testing depending upon whether assumptions about normality and homogeneity of variance were satisfied (van Hecke, 2012). We used three evaluation metrics to assess the performance of seasonal RFs: (1) the area under the curve (AUC) of the receiver operating characteristic (Hanley and McNeil, 1982), (2) the Cohen's Kappa (Kappa; Cohen, 1960), and (3) the true skill statistics (TSS; Allouche et al., 2006). These evaluation metrics have been widely used to test the accuracy of SDMs (Fernandes et al., 2019). Given these metrics, the fitting model was considered acceptable with AUC$\geqslant$0.7, Kappa$\geqslant$0.4 and TSS$\geqslant$0.4 (Phillips et al., 2017; Becker et al., 2020). Optimal parameter settings for different seasonal RFs were determined by the accuracy of values using a cross-validation procedure with 10 evaluation repetitions. Models were run using the biomod2 package in R software version 4.0.2 (Thuiller et al., 2016).

3 Results

3.1 *Model performance and variable importance*

The SDMs of *L. tanakae* were successfully developed for all four seasons. Overall, the seasonal RFs performed well, indicating a highly accurate prediction (mean±standard error) based on the score criteria for the three evaluation metrics. Models for October, August and January were of higher quality compared to that for June.

Among environmental predictors with absolute pairwise Pearson's correlation coefficients less than 0.7, the importance of different variables showed an inconsistent trend across survey month (Fig. 2). In June, bottom temperature was evaluated as the most important variable with the highest importance value of 0.61±0.04, other variables contributed less than 0.20 (Fig. 3a). *Liparis tanakae* prefers to inhabit locations where bottom temperatures ranged from 8°C to 16°C (Fig. 4a). In August and October, bottom salinity was the dominant variable influencing the distribution of *L. tanakae* with importance values of 0.25±0.01 and 0.41±0.02, respectively (Figs 3b and c). The species prefers to inhabit areas with bottom salinity higher than 31.5 (Figs 4b and c). In January, bottom temperature and distance to shore showed higher variable importance than bottom velocity, but all variable importance values were less than 0.20 (Fig. 3d). *Liparis tanakae* prefers to inhabit areas with bottom temperatures higher than 10°C (Fig. 4d). Temperature preference of *L. tanakae* was consistent year-round, ranging from 8°C to 16°C, whereas salinity preference changed slightly from higher than 31.0 in June to higher than 31.5 in August and October.

3.2 *Seasonal habitat suitability*

Habitat suitability for *L. tanakae* showed clear seasonality in the BSYS (Fig. 5). In June, most areas in the BSYS were widely suitable for *L. tanakae* (Fig. 5a). High habitat suitability was predicted in most areas of the Yellow Sea except for coastal waters outside Jiangsu Province. In the Bohai Sea, suitable areas tended to be peripherally distributed except in the central Bohai Sea and three bays (the Bohai Bay, Laizhou Bay and Liaodong Bay). In August and October, *L. tanakae* was more likely to occur in the Yellow Sea, whereas most areas in the Bohai Sea had low probabilities of occurrence, except in the northern area of the Bohai Sea Strait near Dalian (Figs 5b and c). The spatial pattern of suitable habitat receded in October as there was less suitable area in shallow waters compared with August. In January, areas of habitat suitable for *L. tanakae* tended to be narrow compared with the other three survey periods, concentrating in the central and southern Yellow Sea (33°–35°N, 122°–125°E, Fig. 5d).

3.3 *Potential impacts of false absence data*

Potential impacts that false absence data might have on model performance were evaluated by adding different levels of error to the absence data (Fig. 6). Model performance under control and degraded data showed similar patterns for different survey periods. Generally, adding false absence data had strong negative effects on model performance. Evaluation criteria values (AUC, Kappa and TSS) demonstrated an overall downward trend with the added increasing level of errors compared to the control group (data without errors added).

Significance tests showed that the difference in mean AUC between the control group and experimental groups with 30% or more errors added was significant (P<0.05) in June, August, and October. In January, a significant difference was detected between the control group and experimental groups with 20% or

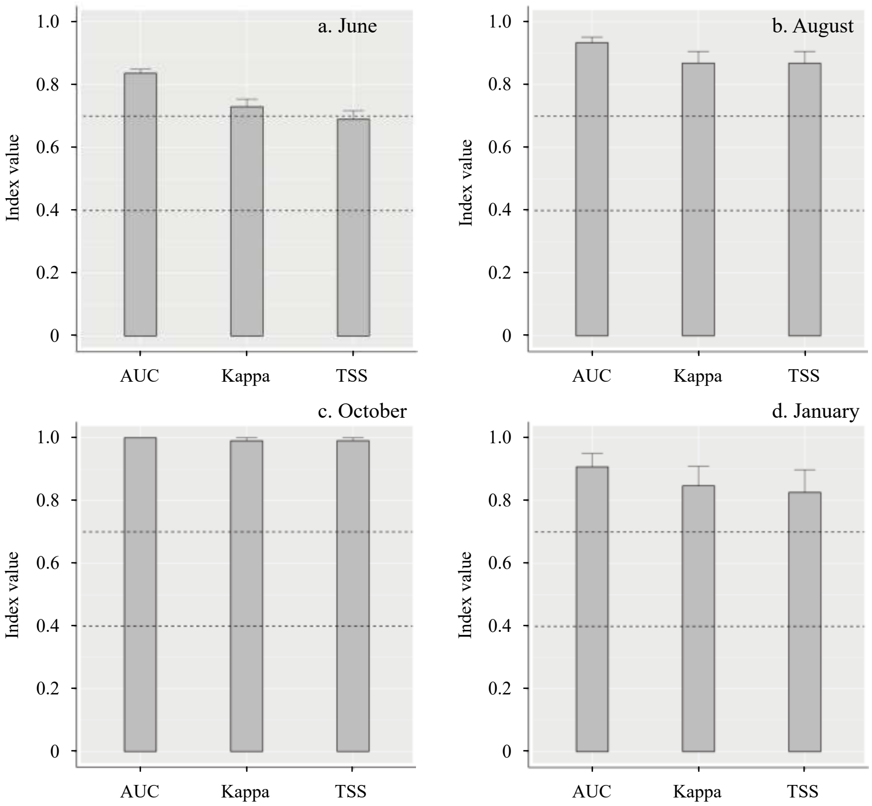

Fig. 2. Model performances evaluated by the area under the receiver operating characteristic curve (AUC), the Cohen's Kappa (Kappa) and the true skill statistics (TSS) of *Liparis tanakae*. Data are expressed with mean±standard error.

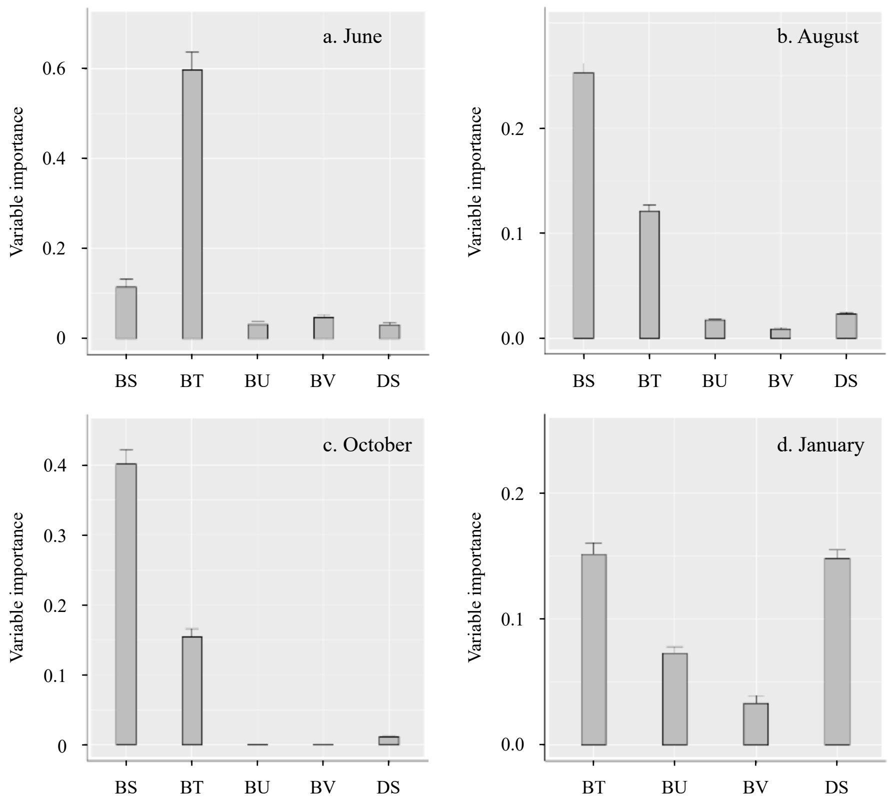

Fig. 3. Seasonal variable importance of environmental variables (BS, bottom salinity; BT, bottom temperature; BU, bottom velocity in latitude direction; BV, bottom velocity in longitude direction; DS, distance to shore). Data are expressed with mean±standard error.

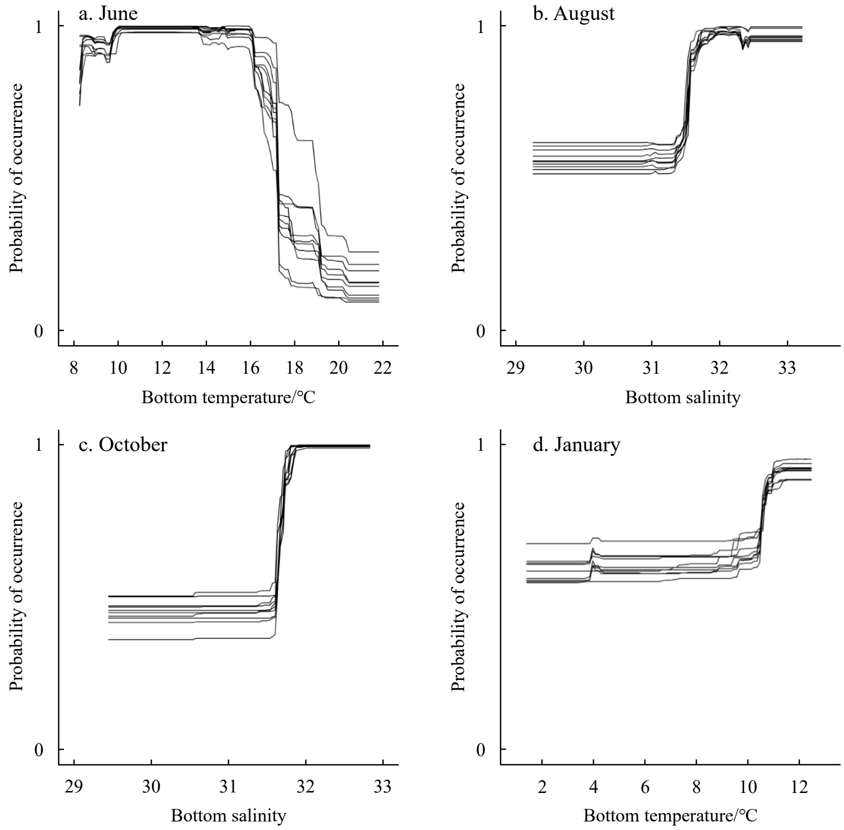

Fig. 4. Response curves of predicted occurrence probability of *Liparis tanakae* against dominant environmental variable. Multiple lines represent the results of 10 evaluation repetitions.

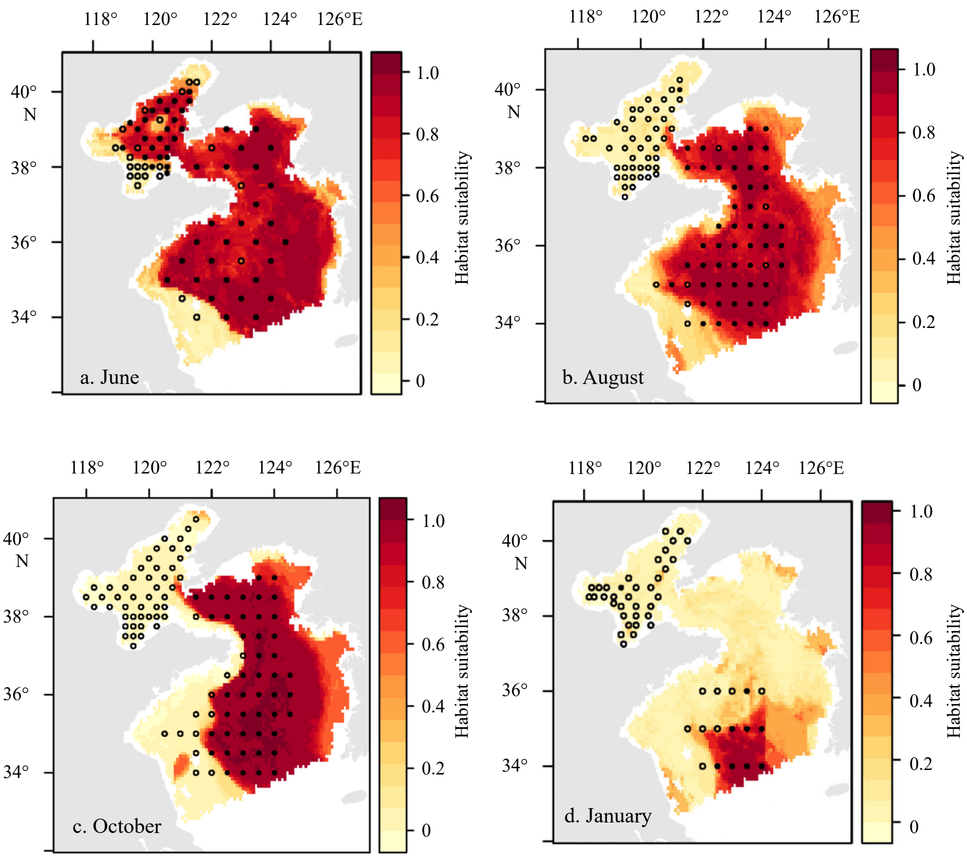

Fig. 5. Seasonal habitat suitability for *L. tanakae* predicted by random forests. Solid dots mean presence; circles, absence.

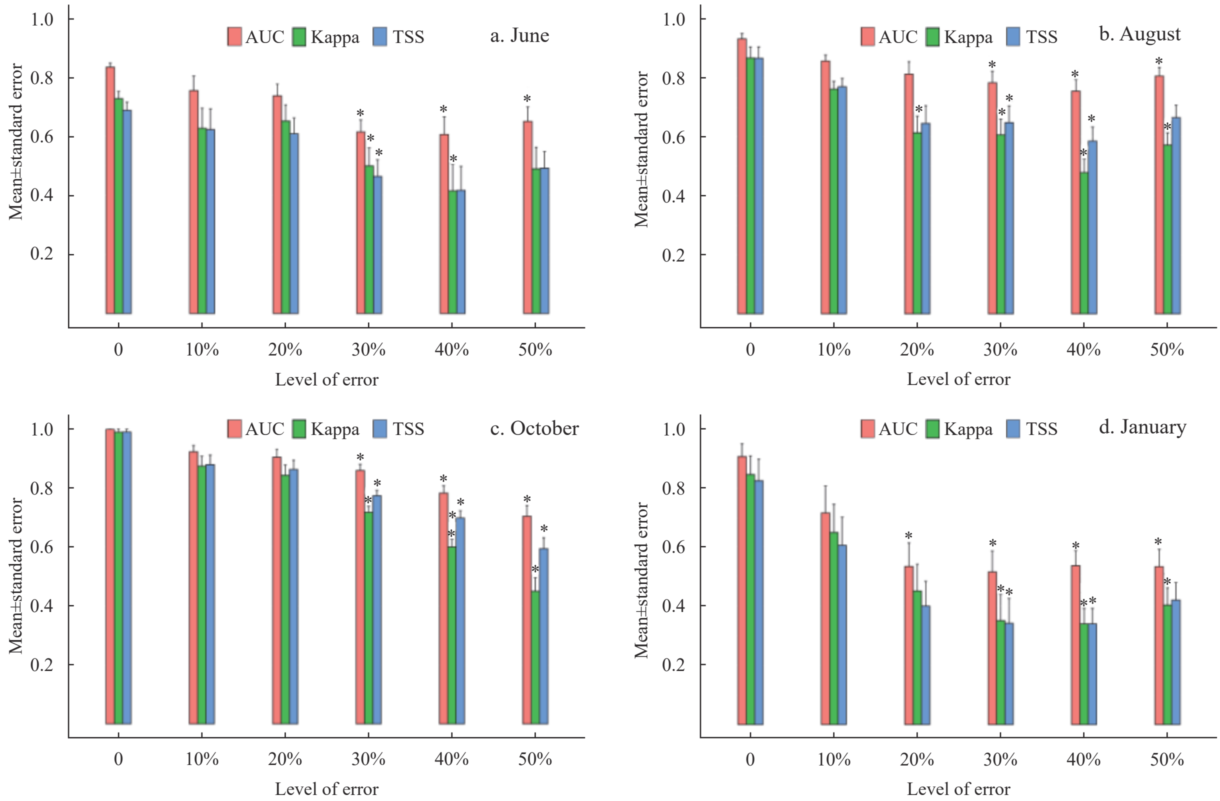

Fig. 6. Model performances of *Liparis tanakae* between control (data without errors added) and degraded data (data with errors added) using RFs. Errors were added to the occurrence dataset to create false negatives. For each level of error, AUC, Kappa and TSS are expressed with mean±standard error.

more errors. Mean Kappa showed a significant difference (P<0.05) between the control group and experimental groups with 30% or higher errors added in January, June and October, whereas in August, that difference was detected after 20% or more errors added. As for the mean value of TSS, it significantly decreased (P<0.05) after 30% or more errors were added in all four survey periods.

4 Discussion

4.1 *Seasonal habitat suitability for L. tanakae*

Accurate species distribution information is necessary for population dynamics assessment, ecosystem modeling and fishery management (Melnychuk et al., 2017; Becker et al., 2020). *Liparis tanakae* has come into prominence as a dominant species of ecological importance in the Yellow Sea (Chen et al., 2018), yet its seasonal distribution pattern remained unclear. To redress this deficiency, we assessed the seasonal habitat suitability for *L. tanakae* in the BYSY for the first time using machine-learning methods. Evaluation values of AUC, Kappa and TSS demonstrated highly accurate prediction among the results. Furthermore, our work revealed that the Yellow Sea appeared to have suitable habitat for *L. tanakae* during all four survey periods. The area of high habitat suitability is extremely wide-ranging, encompassing almost the entire Yellow Sea in all months surveyed except January. Strong environment adaptability may provide a key explanation for the dominant status of *L. tanakae* in the BYSY along with its life history traits such as rapid growth, varied diet and high inter-specific competitiveness (Chen et al., 2013). In addition, our modelling approach should provide greater insight into future changes in spatial distribution patterns of *L. tanakae* when it is applied to further ecosystem modelling work in the study area, especially when seasonal variation is taken into account.

Previous research has indicated that *L. tanakae* inhabits shallower waters during spring and summer compared to autumn and winter (Zhou et al., 2012). This is also reflected in our results in which higher habitat suitability is predicted in coastal waters near the southern region of the Shandong Peninsula in June and August compared to those in October and January. It is noteworthy that the period from May to August is also the spawning season for many economically important fishes and shrimps in the BSYS. During this period, *L. tanakae* belongs to the shrimp predator functional group as well as including a proportion of scale-fish in its diet (Zhang et al., 2011). The occurrence of a large number of *L. tanakae* in shallow waters may have adverse impacts on the recruitment of economically valuable species through predation (Zhou et al., 2012). In the coastal waters off Fukushima, Japan, it is reported that there was a habitat overlap between snailfish (predator) and 0-year-old Japanese flounder (prey) and more than 40% of 0-year-old Japanese flounder were estimated to be vulnerable to *L. tanakae* (Tomiyama et al., 2013b). These ecological issues need to be further explored in the BSYS using methods such as stomach contents analysis and joint species distribution models.

4.2 *Model performance and variable importance*

Environmental variables not only influence the physiological behavior and metabolic activity level of marine species, but also their distribution and population dynamics (Phillips et al., 2017). Considering species presence/absence data and abiotic environmental factors in combination, we aimed at identifying the dominant variables that could explain the seasonal distribution of *L. tanakae*. In our study, bottom temperature and bottom salinity were the most important variables contributing to the distribution patterns of *L. tanakae* in the BSYS, indicating this species prefers relative low temperature and high salinity environments. It has been reported that the preferred habitat of *L. tanakae* is closely related to the distribution of cold coastal water (Chen, 1991). The stock density of *L. tanakae* was negatively correlated with winter sea surface temperature, indicating a sensitive response to changing environmental conditions (Chen et al., 2013). In addition, the dominant variable differed among the survey months. This means that when considering future research involving seasonal population dynamics of *L. tanakae*, the inconsistent role that different environmental parameters may play across seasons should be considered.

Soberón and Nakamura (2009) pointed out that the actual spatial distribution of a species may not be fully explained by the abiotic variables in models, because species dispersal and biotic factors also play important roles and vary with scale in their interactions. In our study, the importance values of the environmental variables were relatively low in January and August, indicating the possible existence of other key variables not included in our study. Spawning and feeding activities of *L. tanakae* in January and August may be related to these lower importance values. It is reported that *L. tanakae* in the Yellow Sea spawns from early January to early March (Wan and Jiang, 2000). There were two intra-annual feeding peaks evident for *L. tanakae*: January, when the feeding intensity was increased to meet reproductive needs, and August, when high feeding intensity supported the demands of rapid growth (Zhang et al., 2011). More abiotic and biological variables should be included in further studies to provide greater insight into factors governing the distribution of *L. tanakae*. In addition, harvest-driven changes in abundance were not considered in our analysis since *L. tanakae* has no economic value and consequently it is not commercially exploited (Chernova et al., 2004).

4.3 *Potential effects of false absence data on model reliability*

Data quality is a key issue influencing the reliability of model predictions (Molloy et al., 2017). High rates of false absences could lead to poorly performing models with low explained deviance (Lobo et al., 2010), especially when a species is rare which makes it difficult to detect all suitable habitat (Gibson et al., 2007). Lobo et al. (2010) defined three types of absences which were contingent, environmental and methodological absences, respectively. Among them, methodological absences were treated as the most important source of uncertainty. In our work it was difficult to distinguish the absence categories due to a lack of more detailed information, so we simply created the “false” absence data by changing absence to presence randomly for a proportion of observations without considering the reasons that caused each absence. Furthermore, it should be noted that prevalence, ecological trait and catchability among species may lead to different degrees of sensitivity to unobserved presence (Comte and Grenouillet, 2013; Manceur and Kühn, 2014). We caution that the projected results for *L. tanakae* in our study are indicative and the negative impacts of false absence need to be checked among multiple species to draw stronger inferences.

In the present study, we tried to evaluate the degree to which false absence data from fishery-independent surveys can impact the accuracy of habitat suitability assessments using different evaluation metrics. Our work revealed that scores for evaluation metrics decreased with an increasing level of error, highlighting the negative effect that the false absence data may have on model performance. Considering the high spatial coverage of the field surveys and the possible level of errors among our absence data, we conclude that our results are acceptable and reliable. It has been established that data quality, modeling methods, sample size and spatial resolution can have large effects on SDMs performance (Molloy et al., 2017; Record et al., 2018; Hao et al., 2019). How these factors affect model performance and their predictive ability in the BSYS are unanswered questions that warrant further investigation.

5 Conclusions

This study utilized seasonal presence and absence information to construct the SDMs for *L. tanakae* in the BSYS. Our results indicated that the seasonal habitat suitability maps of *L. tanakae* were consistent with the field survey data. Seasonal variations should be considered when the spatial distributions of *L. tanakae* are further applied to ecosystem modelling work in the study area. Bottom temperature and salinity were identified as important factors influencing the distribution of *L. tanakae*. The importance values of environmental variables were relatively low during January and August, and biotic variables were not included in our study such as spawning and feeding activities may be related to this pattern. Values of evaluation metrics decreased with an increasing level of error, highlighting the negative effect that false absence data may have on model performance. But we caution that the projected results of *L. tanakae* in our study are indicative and potential negative impacts from false absence, and need to be checked among multiple species. The distributions of *L. tanakae* presented in this paper serve as an example to illustrate the seasonal migration pattern of mobile species in the BSYS.

References

Allouche O, Tsoar A, Kadmon R. 2006. Assessing the accuracy of species distribution models: Prevalence, Kappa and the true skill statistic (TSS). Journal of Applied Ecology, 43(6): 1223–1232, doi: 10.1111/j.1365-2664.2006.01214.x

Barbet-Massin M, Jiguet F, Albert C H, et al. 2012. Selecting pseudo-absences for species distribution models: how, where and how many. Methods in Ecology and Evolution, 3(2): 327–338, doi: 10.1111/j.2041-210X.2011.00172.x

Basher Z, Bowden D A, Costello M J. 2018. Global Marine Environment Datasets (GMED). Version 2.0 (Rev. 02.2018). http://gmed.auckland.ac.nz [2018-07-09/2020-09-19]

Becker L R, Bartholomä A, Singer A, et al. 2020. Small-scale distribution modeling of benthic species in a protected natural hard ground area in the German North Sea (Helgoländer Steingrund). Geo-Marine Letters, 40(2): 167–181, doi: 10.1007/s00367-019-00598-8

Breiman L. 2001. Random forests. Machine Learning, 45(1): 5–32, doi: 10.1023/A:1010933404324

Brotons L, Thuiller W, Araújo M B, et al. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography, 27(4): 437–448, doi: 10.1111/j.0906-7590.2004.03764.x

Chen Dagang. 1991. Fishery Ecology of the Bohai Sea and the Yellow Sea (in Chinese). Beijing: China Ocean Press, 383–386

Chen Yunlong, Shan Xiujuan, Jin Xianshi, et al. 2018. Changes in fish diversity and community structure in the central and southern Yellow Sea from 2003 to 2015. Journal of Oceanology and Limnology, 36(3): 805–817, doi: 10.1007/s00343-018-6287-6

Chen Yunlong, Shan Xiujuan, Ovando D, et al. 2021. Predicting current and future global distribution of black rockfish (*Sebastes schlegelii*) under changing climate. Ecological Indicators, 128: 107799, doi: 10.1016/j.ecolind.2021.107799

Chen Yunlong, Shan Xiujuan, Zhou Zhipeng, et al. 2013. Interannual variation in the population dynamics of snailfish *Liparis tanakae* in the Yellow Sea. Acta Ecologica Sinica, 33(19): 6227–6235, doi: 10.5846/stxb201304170731

Chernova N V, Stein D L, Andriashev A P. 2004. Family Liparidae Scopoli 1777—snailfishes. In: Annotated Check lists of Fishes. No. 31. San Francisco: California Academy of Sciences, 1–72

Cohen J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1): 37–46, doi: 10.1177/001316446002000104

Comte L, Grenouillet G. 2013. Species distribution modelling and imperfect detection: comparing occupancy versus consensus methods. Diversity and Distributions, 19(8): 996–1007, doi: 10.1111/ddi.12078

Cutler D R, Edwards Jr T C J, Beard K H, et al. 2007. Random forests for classification in ecology. Ecology, 88(11): 2783–2792, doi: 10.1890/07-0539.1

Elith J, Leathwick J R, Hastie T. 2008. A working guide to boosted regression trees. Journal of Animal Ecology, 77(4): 802–813, doi: 10.1111/j.1365-2656.2008.01390.x

Fernandes R F, Scherrer D, Guisan A. 2019. Effects of simulated observation errors on the performance of species distribution models. Diversity and Distributions, 25(3): 400–413, doi: 10.1111/ddi.12868

Fu Caihong, Olsen N, Taylor N, et al. 2017. Spatial and temporal dynamics of predator-prey species interactions off western Canada. ICES Journal of Marine Science, 74(8): 2107–2119, doi: 10.1093/icesjms/fsx056

Gibson L, Barrett B, Burbidge A. 2007. Dealing with uncertain absences in habitat modelling: a case study of a rare ground-dwelling parrot. Diversity and Distributions, 13(6): 704–713, doi: 10.1111/j.1472-4642.2007.00365.x

Guo Yanning, Xu Zhen, Zhang Luping, et al. 2014. Occurrence of *Hysterothylacium* and *Anisakis* nematodes (Ascaridida: Ascaridoidea) in the tanaka's snailfish *Liparis tanakae* (Gilbert & Burke) (Scorpaeniformes: Liparidae). Parasitology Research, 113(4): 1289–1300, doi: 10.1007/s00436-014-3767-2

Hanley J A, McNeil B J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143(1): 29–36, doi: 10.1148/radiology.143.1.7063747

Hao Tianxiao, Elith J, Guillera-Arroita G, et al. 2019. A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. Diversity and Distributions, 25(5): 839–852, doi: 10.1111/ddi.12892

Hecke T V. 2012. Power study of anova versus Kruskal-Wallis test. Journal of Statistics and Management Systems, 15(2–3): 241–247, doi: 10.1080/09720510.2012.10701623

Jin Xianshi, Tang Qisheng. 1996. Changes in fish species diversity and dominant species composition in the Yellow Sea. Fisheries Research, 26(3–4): 337–352, doi: 10.1016/0165-7836(95)00422-X

Jin Xianshi, Xu Binduo, Tang Qisheng. 2003. Fish assemblage structure in the East China Sea and southern Yellow Sea during autumn and spring. Journal of Fish Biology, 62(5): 1194–1205, doi: 10.1046/j.1095-8649.2003.00116.x

Jin Xianshi, Zhang Bo, Xue Ying. 2010. The response of the diets of four carnivorous fishes to variations in the Yellow Sea ecosystem. Deep-Sea Research Part II: Topical Studies in Oceanography, 57(11–12): 996–1000, doi: 10.1016/j.dsr2.2010.02.001

Lauria V, Gristina M, Attrill M J, et al. 2015. Predictive habitat suitability models to aid conservation of elasmobranch diversity in the central Mediterranean Sea. Scientific Reports, 5: 13245, doi: 10.1038/srep13245

Lobo J M, Jiménez-Valverde A, Hortal J. 2010. The uncertain nature of absences and their importance in species distribution modelling. Ecography, 33: 103–114, doi: 10.1111/j.1600-0587.2009.06039.x

Manceur A M, Kühn I. 2014. Inferring model-based probability of occurrence from preferentially sampled data with uncertain absences using expert knowledge. Methods in Ecology and Evolution, 5(8): 739–750, doi: 10.1111/2041-210x.12224

Marx M, Quillfeldt P. 2018. Species distribution models of European Turtle Doves in Germany are more reliable with presence only rather than presence absence data. Scientific Reports, 8(1): 16898, doi: 10.1038/s41598-018-35318-2

Melnychuk M C, Peterson E, Elliott M, et al. 2017. Fisheries management impacts on target species status. Proceedings of the National Academy of Sciences of the United States of America, 114(1): 178–183, doi: 10.1073/pnas.1609915114

Molloy S W, Davis R A, Dunlop J A, et al. 2017. Applying surrogate species presences to correct sample bias in species distribution models: a case study using the Pilbara population of the Northern Quoll. Nature Conservation, 18: 27–46, doi: 10.3897/natureconservation.18.12235

Park J M, Kwak S N, Huh S H, et al. 2017. Diets and niche overlap among nine co-occurring demersal fishes in the southern continental shelf of East/Japan Sea, Korea. Deep-Sea Research Part II: Topical Studies in Oceanography, 143: 100–109, doi: 10.1016/j.dsr2.2017.06.002

Phillips N D, Reid N, Thys T, et al. 2017. Applying species distribution modelling to a data poor, pelagic fish complex: the ocean sunfishes. Journal of Biogeography, 44(10): 2176–2187, doi: 10.1111/jbi.13033

Pons M, Melnychuk M C, Hilborn R. 2018. Management effectiveness of large pelagic fisheries in the high seas. Fish and Fisheries, 19(2): 260–270, doi: 10.1111/faf.12253

Record S, Strecker A, Tuanmu M N, et al. 2018. Does scale matter? A systematic review of incorporating biological realism when predicting changes in species distributions. PLoS ONE, 13(4): e0194650, doi: 10.1371/journal.pone.0194650

Rubio I, Ganzedo U, Hobday A J, et al. 2020. Southward re-distribution of tropical tuna fisheries activity can be explained by technological and management change. Fish and Fisheries, 21(3): 511–521, doi: 10.1111/faf.12443

Sarquis J A, Cristaldi M A, Arzamendia V, et al. 2018. Species distribution models and empirical test: comparing predictions with well-understood geographical distribution of *Bothrops alternatus* in Argentina. Ecology and Evolution, 8(21): 10497–10509, doi: 10.1002/ece3.4517

Schickele A, Leroy B, Beaugrand G, et al. 2020. Modelling European small pelagic fish distribution: Methodological insights. Ecological Modelling, 416: 108902, doi: 10.1016/j.ecolmodel.2019.108902

Soberón J, Nakamura M. 2009. Niches and distributional areas: concepts, methods, and assumptions. Proceedings of the National Academy of Sciences of the United States of America, 106(S2): 19644–19650, doi: 10.1073/pnas.0901637106

Tanaka K R, Torre M P, Saba V S, et al. 2020. An ensemble high-resolution projection of changes in the future habitat of American lobster and sea scallop in the Northeast US continental shelf. Diversity and Distributions, 26(7): 987–1001, doi: 10.1111/ddi.13069

Thuiller W, Georges D, Gueguen M, et al. 2016. Biomod2: Ensemble platform for species distribution modeling. https://cran.r-project.org/package=biomod2 [2021-06-11/2021-07-22]

Tomiyama T, Uehara S, Kurita Y. 2013a. Feeding relationships among fishes in shallow sandy areas in relation to stocking of Japanese flounder. Marine Ecology Progress Series, 479: 163–175, doi: 10.3354/meps10191

Tomiyama T, Yamada M, Yoshida T. 2013b. Seasonal migration of the snailfish *Liparis tanakae* and their habitat overlap with 0-year-old Japanese flounder *Paralichthys olivaceus*. Journal of the Marine Biological Association of the United Kingdom, 93(7): 1981–1987, doi: 10.1017/S0025315413000544

Wan Ruijing, Jiang Yanwei. 2000. The species and biological characteristics of the eggs and larvae of osteichthyes in the Bohai Sea and Yellow Sea. Journal of Shanghai Fisheries University, 9(4): 290–297

Wang Fan, Liu Chuanyu. 2009. An N-shape thermal front in the western South Yellow Sea in winter. Chinese Journal of Oceanology and Limnology, 27(4): 898, doi: 10.1007/s00343-009-9045-y

Wisz M S, Broennimann O, Grønkjær P, et al. 2015. Arctic warming will promote Atlantic-Pacific fish interchange. Nature Climate

Change, 5(3): 261–265, doi: 10.1038/nclimate2500
Zhang Bo, Jin Xianshi, Dai Fangqun. 2011. Feeding habits and their variation of seasnail (*Liparis tanakae*) in the central and southern Yellow Sea. Journal of Fisheries of China, 35(8): 1199–1207
Zhong Mingyu, Wu Huifeng, Mi Wenying, et al. 2018. Occurrences and distribution characteristics of organophosphate ester flame retardants and plasticizers in the sediments of the Bohai and Yellow Seas, China. Science of The Total Environment, 615: 1305–1311, doi: 10.1016/j.scitotenv.2017.09.272
Zhou Zhipeng, Jin Xianshi, Shan Xiujuan, et al. 2012. Seasonal variations in distribution and biological characteristics of snailfish *Liparis tanakae* in the central and southern Yellow Sea. Acta Ecologica Sinica, 32(17): 5550–5561, doi: 10.5846/stxb201108061152

Supplementary information:

Fig. S1. Results of monthly mean surface and bottom salinity from the ROMS model and Hybrid Coordinate Ocean Model (HYCOM) in 2016.

Fig. S2. Results of monthly mean sea surface temperature from the NOAA optimum interpolation sea surface temperature (OISST), ROMS model and Hybrid Coordinate Ocean Model (HYCOM) in 2016.

Fig. S3. Simulations of monthly mean temperature and salinity in July from the ROMS model.

Fig. S4. Correlation coefficients matrix of five environmental variables.

Text S1. Validation of the oceanographic data.

The supplementary information is available online at https://doi.org/10.1007/s13131-021-1912-1 and www.aosocean.com. The supplementary information is published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.