Species Distribution Model Predictions of the Critically Endangered Grey Nurse Shark in Australia

Guanfang Su (✉ bi1996@163.com)

Jilin University

Research Article

Keywords: shark, models, model, species, used, grey, nurse

Posted Date: July 14th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-677819/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Species distribution model predictions of the Critically Endangered grey nurse
shark in Australia**

Guanfang Su ^{1*}

¹Affiliation:

¹Email: bi1996@163.com

Abstract

Species distribution models (SDMs) are commonly used to forecast how threatened species are influenced by climate change. The grey nurse shark (*Carcharias taurus*) is a critically endangered species inhabiting both the east and west coasts of Australia, with negligible genetic interchange between the two populations. I used Generalized Linear Models (GLM), Maximum Entropy (MaxEnt) models and Boosted Regression Trees (BRT) to predict the distribution of the grey nurse shark. The data were a sample of presence-only data, derived from the known grey nurse shark sighting locations, from the east coasts of Australia, with pseudo-absences generated and bootstrapped from a restricted background. I verified these models using leave-one-out cross validation and model metrics including AICc, BIC, percentage of deviance explained, leave-one-out cross-validated R^2 , AUC, maximum Cohen's Kappa, specificity and sensitivity. Cross-validated R^2 was used as an overall comparison method across model types. I performed out-of-source validation by comparing model projection with the distributional range of the ragged tooth shark (*Carcharias taurus*) in South Africa. The prediction of the selected model was consistent with the current distributional range of the ragged tooth shark.

Introduction

Species distribution modelling has been widely used in conservation biology and ecology ¹ for discovering new species ², prediction of distributional ranges ^{3,4}, modelling invasive species' potential distribution ⁵, and detecting climate impacts on the habitat suitability ⁶. The species distribution model has been used in many domains, in combination with ecological niche theory, at both individual and community levels ⁷⁻¹⁰.

These distribution models use different mechanisms to infer the characteristics of the preferred habitat of target species ¹¹. Many methods have been subjected to model comparison and performance evaluation, with no obviously best method that is suitable for all situations ¹². Model accuracy is difficult to quantify because the evaluation methods (threshold-dependent and threshold-independent criteria) are susceptible and there is typically a lack of sufficient data to do independent, out-of-sample performance tests ^{13,14}.

The selection of pseudo-absence data can also exert a strong influence on subsequent predictions ¹⁵. For instance, the questions of how large a background region to cover, and how many pseudo-absence points are optimal, are not settled issues. Some studies have shown that the relative importance of predictors and performance statistics such are sensitive to scale and background ^{16,17}. However, the relationship between species response curve fitting and pseudo-absence data selection remains an inadequately resolved problem ⁸.

As an alternative to making assumptions about pseudo-absences, presence-only models such as MaxEnt ¹⁸, GARP ¹⁹, and ENFA ²⁰ can be used when no absence data is available. They detect the marginality of the species in the entire background ²¹. Again, however, the question of how background scale and frequency distribution influence the model performance to the entire environment is not well studied.

I developed a case study of the grey nurse shark (*Carcharias taurus*, Rafinesque 1810), a coastal marine predatory fish with a small presence sample size and very wide distribution along the coast of Australia, to test the abiotic factor impact on this top marine predator's distribution. In Australia, several broad-scale surveys in the states of New South Wales, Queensland and Western Australia have been done to identify the species' aggregation sites and migration patterns; however, its west-coast distribution is poorly quantified and only one aggregation site was identified ²². Historically, the east and west coast populations have been separated and have negligible genetic interchange ²³. It has been argued that the model with better transferability is better algorithm ²⁴. As such, I used the east coast sighting locations for model fitting and projected the best fitted model in South African coast for model testing. Presence-absence and Presence-only models GLM, BRT and MaxEnt were fitted.

Methods

Species data

I collated data describing the major sighting sites of the east coast population of grey nurse sharks from the *Issues Paper for the Grey Nurse Shark (Carcharias taurus)* generated by Australian Government

(<https://environment.gov.au/system/files/resources/91e141d0-47aa-48c5-8a0f-992b9df960fe/files/issues-paper-grey-nurse-shark-carcharias-taurus.pdf>). The original presence data consisted of 153 points. However, there were 17 points falling on the grids where the elevation values were positive. I deleted these data points from the presence data. There was one presence point lying on the grid where the bathymetric slope value was zero. I removed this point in order to apply logarithm transformation to the bathymetric slope data. Thus, the number of the input presence points was 135.

Environmental data

The four prime predictors were bathymetric slope, depth, 15-year January Sea Surface Temperature (SST) mean and 15-year January interannual SST variance. The Australian Bathymetry and Topography Grid dataset was used for model fitting. The resolution of this dataset is 0.0025°. The ETOPO1 1 Arc-Minute Global Relief Model was obtained for model projection in South African coastal area. DEM Surface Tools in ArcGIS 10 (ESRI 2011. ArcGIS Desktop: Release 10. Redlands, California, USA) was used to calculate the bathymetric slope data set. The SST mean and variance

predictors were generated from the 4 km Advanced Very High Resolution Radiometer (AVHRR) Pathfinder Version 5.0. The SST mean and variance were bilinearly interpolated, so that the resolutions of the four predictors were the same.

Logarithm transformation was applied to bathymetric slope and January SST mean. The square of the log-transformed January SST mean was calculated for GLM and BRT. I scaled bathymetric depth and interannual SST variance according to the following equation (1):

$$X_{ij}^n = \frac{X_{ij} - X_{j,min}}{X_{j,max} - X_{j,min}} \quad (1)$$

where X_{ij}^n and X_{ij} are the scaled and non-scaled j th variable values of point i .

$X_{j,max}$ and $X_{j,min}$ are the maximum and minimum values of j th variable. Because MaxEnt applies this scaling to all its predictors automatically, the scaled SST variance, log-transformed bathymetric slope, non-transformed SST mean and non-scaled bathymetric depth were used as input for MaxEnt models.

Model types

I employed three prevalent species distribution models: GLM²⁵, MaxEnt¹⁸ and BRT^{26,27}. Generalized linear models are widely used in species distribution modelling^{28,29}.

GLM is generalized multiple variates regression analysis using the link function³⁰. I used binomial distribution with a logit link function to fit the presence and pseudo-absence dataset. MaxEnt (version 3.3.3k, AT&T Labs Research) is a presence-only modelling technique using both background and presence data. MaxEnt fits

coefficients by minimizing the regularized negative log-likelihood to achieve the maximum entropy of the predicted distribution. I used linear and quadratic features for MaxEnt model fitting. Regression Tree model makes use of least-square regression to estimate parameters ²⁷. The interaction depth was set to 3 to avoid overfitting, and the bag fraction was set to 1. The shrinkage value and the number of trees were set to 0.01 and 200, respectively.

Pseudo-absence point generation

Grey nurse sharks are mainly found in the shallow neritic waters, so a common sample bias is always detected in the data ^{31,32}. I used the restricted background method to reduce this sample bias ³³. I restricted the maximum depth of background layers to 93 m, for the maximum depth of the presence points was 89 m. I extended the range of the background layers to 1° further than that of the presence points ^{17,33-35} (see Figure 1). I also eliminated the background data points lying on the grids where the bathymetric slope values were zero. The number of pseudo-absence points in each data set was set to 1350, which was 10 times as big as the number of presence points.

I bootstrapped the pseudo-absences from the restricted background using randomly sampling method with replacement in R statistical software (R Core Team). I generated 500 bootstrap pseudo-absence samples to form 500 data sets. Each data set had 135 presences and 1350 pseudo-absences. I then randomly split each data set

into 135 training data sets and 135 testing data sets to achieve leave-one-out cross validation.

Model evaluation and out-of-source validation

Various model metrics were used for model evaluation. Firstly, the sample size corrected Akaike information criterion (AIC) and Bayesian Information Criterion (BIC) ³⁶ were calculated for GLM and MaxEnt models ³⁷. Secondly, leave-one-out cross-validated coefficient of determination (cross-validated R^2) ³⁸ were computed for measuring the goodness-of-fit of the candidate models. Cross-validated R^2 was used as a universal model evaluation method across different model types. For MaxEnt, the logistic output values were used to calculate the cross-validated R^2 . Thirdly, threshold-independent Area Under the Curve (AUC) ³⁹, maximum Cohen's Kappa ⁴⁰, specificity and sensitivity ^{37,39,41} were also calculated by leave-one-out cross validation. I then estimated the median, 5th and 95th percentiles for each model metric ⁴².

I projected the median cross-validated R^2 selected candidate model in the coastal area of South Africa as an out-of-source validation. In this validation, the bathymetric depth was restricted and the points where bathymetric slope values were zero were eliminated. All the variables in the out-of-source dataset were transformed and scaled according to the methods mentioned before.

Results

Model evaluation

I used cross-validated R^2 , which is asymptotically equivalent to AIC-type criteria, as a universal method for model selection across the three model types. The model ranking results were shown in Table 1, Table 2 and Table 3.

According to the median values of these model metrics, one of the GLM candidate models gained the highest median value of cross-validated R^2 .

This model comprised bathymetric slope, depth, January SST mean and January SST mean square as predictors. The frequencies that this model was selected among GLM candidate models by AICc, BIC and cross-validated R^2 were 72.7%, 99.8% and 77.6%. The median value of cross-validated R^2 of this model was 0.3569, which was the highest across three model types. The median values of AUC, Maximum Cohen's Kappa and the sum of Sensitivity and Specificity of this model were also the highest across three model types.

With respect to MaxEnt models, the cross-validated R^2 of some candidate models were negative. The ranking results by BIC and cross-validated R^2 for MaxEnt candidate models were the same. The MaxEnt candidate model selected by AICc was the full model while BIC and cross-validated R^2 selected one was comprised of bathymetric slope, depth, January SST mean and the square of bathymetric slope, the square of bathymetric depth and the square of January SST mean as predictors. The frequencies that the full model was selected as the best model by AICc, BIC and

cross-validated R^2 were 59.8%, 19.7% and 39% while as for the BIC selected model, the frequencies were 40.2%, 78.9% and 61%. The Sensitivity values of MaxEnt models were higher than those of GLM and BRT models.

The cross-validated R^2 selected BRT candidate model consisted of bathymetric slope, depth, January SST mean and January SST mean square as dependent variables. The median value of cross-validated R^2 for this model was 0.3058. The median value of AUC gained by this model was 0.9094, which was lower than that of the best GLM model.

Out-of-source validation

I projected the GLM candidate model with the highest median value of cross-validated R^2 in the South African and Australian coastal areas. Model projections were consistent with the current distributional ranges of the ragged tooth shark (*Carcharias taurus*) and the grey nurse shark (*Carcharias taurus*). Figure 2 and Figure 3 were averaged predictions from 500 simulations. The distributional range of ragged tooth sharks in the coast of South Africa was from around Sodwana Bay, Kwazulu-Natal to Cape Town ⁴³. As for the east coast population in Australia, the sighting locations of the grey nurse shark ranged from Fraser Island, Queensland to Gabo Island, Victoria (*Issue paper for the grey nurse shark (Carcharias taurus)*). It is reported in the *Issue paper* that the range of the west coast population in Australia was from Exmouth, Western Australia to the coastal area near Cocklebidy in the

Great Australian Bight. The model predictions were in line with these ranges (see Figure 2 and Figure 3).

Discussion

The predicted habitat suitability was mainly dependent on the multivariate response surface⁴⁴. Carefully examining the multivariate response surface is important. The response curve of January SST mean of the cross-validated R^2 selected GLM model was a narrow bell-shaped curve, which may resemble the realised niche of the species. Grey nurse sharks' aggregation sites are always found around inshore islands and rocky reefs with pinnacles, gutters and rocky caves. Thus, the bathymetric slope might be used as a surrogate for depicting the complex sea bed structure that the shark prefers. As for the bathymetric depth, the model indicated a trend that shallow waters were more preferable than deep waters for the species. However, although the predicted habitat suitability in waters deeper than 90 m was close to zero, the maximum depth to which the shark dives is deeper than 200 m, indicating a sampling bias influenced the model performance.

Various model evaluation methods were explored in this study. Overall, although the median value of cross-validated R^2 of the best fitted GLM model was not very high, the model's projections were consistent with the current distributional ranges of this species in South Africa and Australia, indicating that it captured the inherent responses to the environmental gradients⁸. From the model selection results, it was

easy to observe that the SST mean, bathymetric slope and depth were dominant variables. In this study, the interannual SST variance was highly correlated with the SST mean, which may result in model selection bias.

Data availability

The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

References

- 1 Carvalho, S. B., Brito, J. C., Pressey, R. L., Crespo, E. & Possingham, H. P. Simulating the effects of using different types of species distribution data in reserve selection. *Biological Conservation* **143**, 426-438 (2010).
- 2 Raxworthy, C. *et al.* Predicting distributions of known and unknown reptile species in Madagascar. *Nature* **426**, 837-841, doi:10.1038/nature02205 (2004).
- 3 Guisan, A., Theurillat, J. P. & Kienast, F. Predicting the potential distribution of plant species in an alpine environment. *Journal of Vegetation Science* **9**, 65-74, doi:10.2307/3237224 (1998).
- 4 Prates-Clark, C. D. C., Saatchi, S. S. & Agosti, D. Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data. *Ecological Modelling* **211**, 309-323, doi:10.1016/j.ecolmodel.2007.09.024 (2008).
- 5 Hortal, J., Borges, P., Jiménez-Valverde, A., Azevedo, E. & Silva, L. Assessing the areas under risk of invasion within islands through potential distribution modelling: The case of *Pittosporum undulatum* in Sao Miguel, Azores. *Journal for Nature Conservation* **18**, 247-257, doi:10.1016/j.jnc.2009.11.002 (2010).
- 6 Dormann, C. F. Promising the future? Global change projections of species distributions. *Basic & Applied Ecology* **8**, 387-397 (2007).
- 7 Olden, J., Joy, M. & Death, R. Rediscovering the species in community-wide predictive modeling. *Ecological applications* **16**, 1449-1460, doi:10.1890/1051-0761(2006)016[1449:RTSICP]2.0.CO;2 (2006).

- 8 Austin, M. Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling* **200**, 1-19, doi:10.1016/j.ecolmodel.2006.07.005 (2007).
- 9 Hirzel, A. & Le Lay, G. Habitat suitability modelling and niche theory. *Journal of Applied Ecology* **45**, 1372-1381, doi:10.1111/j.1365-2664.2008.01524.x (2008).
- 10 Braunisch, V., Bollmann, K., Graf, R. F. & Hirzel, A. H. Living on the edge—modelling habitat suitability for species at the edge of their fundamental niche. *Ecological Modelling* **214**, 153-167 (2008).
- 11 Guisan, A. & Zimmermann, N. Predictive habitat distribution models in ecology. *Ecological Modeling. Ecological Modelling* **135**, 147-186, doi:10.1016/S0304-3800(00)00354-9 (2000).
- 12 Elith, J. & Graham, C. H. Do they? How do they? Why do they? On finding reasons for differing performances of species distribution models. *Ecography* **32**, 66-77 (2009).
- 13 Moisen, G. G., Edwards, T. C. & Osborne, P. E. Further advances in predicting species distributions. *Ecological Modelling* **199**, 129-131, doi:10.1016/j.ecolmodel.2006.05.015 (2006).
- 14 Peterson, A., Papes, M. & Soberón, J. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* **213**, 63-72, doi:10.1016/j.ecolmodel.2007.11.008 (2008).
- 15 Chefaoui, R. M. & Lobo, J. M. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling* **210**, 478-486 (2008).
- 16 Santika, T. & Hutchinson, M. F. The effect of species response form on species distribution model prediction and inference. *Ecological Modelling* **220**, 2365-2379, doi:10.1016/j.ecolmodel.2009.06.004 (2009).
- 17 Vanderwal, J., P. Shoo, L., Graham, C. & Williams, S. Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modelling* **220**, 589-594, doi:10.1016/j.ecolmodel.2008.11.010 (2009).
- 18 Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**, 231-259 (2006).
- 19 Stockwell, D. & Peters, D. The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science* **13**, 143-158, doi:10.1080/136588199241391 (1999).
- 20 Hirzel, A., Le Lay, G., Helfer, V., Randin, C. & Guisan, A. Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling* **199**, 142-152, doi:10.1016/j.ecolmodel.2006.05.017 (2006).

- 21 Hirzel, A., Hausser, J., Chessel, D. & Perrin, N. Ecological-Niche Factor Analysis: How to Compute Habitat-Suitability Maps without Absence Data? *Ecology* **83**, 2027-2036, doi:10.2307/3071784 (2002).
- 22 Hoschke, A. M. & Whisson, G. J. First aggregation of grey nurse sharks (*Carcharias taurus*) confirmed in Western Australia. *Marine Biodiversity Records* **9**, 1-10 (2016).
- 23 Stow, A. *et al.* Isolation and genetic diversity of endangered grey nurse shark (*Carcharias taurus*) populations. *Biology letters* **2**, 308-311, doi:10.1098/rsbl.2006.0441 (2006).
- 24 Pearson, R. G., Raxworthy, C., Nakamura, M. & Peterson, A. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* **34**, 102-117, doi:10.1111/j.1365-2699.2006.01594.x (2007).
- 25 Nelder, J. A. & Wedderburn, R. W. M. Generalised Linear Model. *Journal of the Royal Statistical Society* **135**, 370-384, doi:10.2307/2344614 (1972).
- 26 Friedman, J., Hastie, T. & Tibshirani, R. Additive logistic regression: a statistical view of boosting: rejoinder. *The Annals of Statistics* **28**, 400-407, doi:10.1214/aos/1016218223 (2000).
- 27 Elith, J., Leathwick, J. R. & Hastie, T. A working guide to Boosted Regression Trees. *The Journal of animal ecology* **77**, 802-813, doi:10.1111/j.1365-2656.2008.01390.x (2008).
- 28 Guisan, A., Weiss, S. & D. Weiss, A. GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology* **143**, 107-122, doi:10.1023/A:1009841519580 (1999).
- 29 Guisan, A., C Edwards, T. & Hastie, T. Generalized Linear and Generalized Additive Models in studies of species distributions: setting the scene. *Ecological Modelling* **157**, 89-100, doi:10.1016/S0304-3800(02)00204-1 (2002).
- 30 Hirzel, A., Helfer, V. & Metral, F. Assessing habitat-suitability models with a virtual species. *Ecological Modelling* **145**, 111-121, doi:10.1016/S0304-3800(01)00396-9 (2001).
- 31 Gilmore, R. G. Reproductive biology of lamnoid sharks. *Environmental Biology of Fishes* **38**, 95-114, doi:10.1007/BF00842907 (1993).
- 32 Otway, N. M., Burke, A. L., Morrison, N. S. & Parker, P. C. Monitoring and identification of NSW critical habitat sites for conservation of grey nurse sharks. 62 (NSW Fisheries Office of Conservation, Nelson Bay, NSW, 2003).
- 33 Phillips, S. J. *et al.* Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological applications* **19**, 181-197, doi:10.1890/07-2153.1 (2009).
- 34 Austin, M. P. & Meyers, J. A. Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. *Forest Ecology & Management* **85**, 95-106 (1996).

- 35 Thuiller, W., Brotons, L., Araújo, M. & Lavorel, S. Effects of restricting range of data to project current and future species distributions. *Ecography* **27**, 165-172, doi:10.1111/j.0906-7590.2004.03673.x (2004).
- 36 Burnham, K. P. & Anderson, D. R. *Model selection and multi-model inference: a practical information-theoretical approach*. (Springer, 2002).
- 37 Warren, D. & Seifert, S. N. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* **21**, 335-342 (2011).
- 38 Efron, B. Regression and ANOVA with zero-one data: measures of residual variation. *Journal of the American Statistical Association* **73**, 113-121 (1978).
- 39 Hanley, J. A. & McNeil, B. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. *Radiology* **143**, 29-36, doi:10.1148/radiology.143.1.7063747 (1982).
- 40 Cohen, J. A. A coefficient of agreement for nominal scales. *Educational & Psychological Measurement* **20**, 37-46, doi:10.1177/0013164460020000104 (1960).
- 41 Ringham, B., Alonzo, T., Grunwald, G. & Glueck, D. Estimates of sensitivity and specificity can be biased when reporting the results of the second test in a screening trial conducted in series. *BMC medical research methodology* **10**, 3, doi:10.1186/1471-2288-10-3 (2010).
- 42 Elith, J. & Leathwick, J. R. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology Evolution & Systematics* **40**, 677-697 (2009).
- 43 Dicken, M., Smale, M. & Booth, A. Spatial and seasonal distribution patterns of the ragged-tooth shark *Carcharias taurus* along the coast of South Africa. *South African Journal of Marine Science* **28**, 603-616 (2006).
- 44 Elith, J., Ferrier, S., Huettmann, F. & Leathwick, J. R. The evaluation strip: A new and robust method for plotting predicted responses from species distribution models. *Ecological Modelling* **186**, 280-289, doi:10.1016/j.ecolmodel.2004.12.007 (2005).

Author contributions

G. S. designed the case study, collected the data, and did the analysis. The first draft of the manuscript was written and then revised by G. S.

Competing interests

The author declares no competing interests.

Fig 1 Sighting locations and restricted background in the east coast of Australia

Fig 2 The projection of the best GLM candidate model in South African coastal area (the resolution of environmental layers was interpolated into 0.04394°)

Fig 3 The projection of the best GLM candidate model in the coasts of Australia

(the resolution of environmental layers was aggregated from 0.0025° to 0.1°)

Table 1 Median values of model metrics (Slope.l: log-transformed bottom slope; JanSSTm.l: log-transformed January SST mean; JanSSTm.l²: the square of the log-transformed January SST mean; JanSSTvar.c: scaled interannual variance of January SST; Depth.c: scaled bathymetric depth; JanSSTm: January SST mean; Depth: bathymetric depth; LogL: log-likelihood; AICc: a sample size corrected AIC; dAICc: the difference between AIC value of a candidate model and the smallest AIC value in the set of candidate models; wAICc: AICc weight; BIC: Bayesian Information Criterion; dBIC: the difference between BIC value of a candidate model and the smallest BIC value in the set of candidate models; wBIC: BIC weight; BIC differences; pcdev: percentage of explained deviance; cvweight: CV weight; r2n_pred: cross-validated R²)

GLM (median)	k	LogL	AICc	BIC	dAICc	wAICc	dBIC	wBIC	pcdev	r2n_pred	AUC	Kappa	specificity	sensitivity
PA ~ Slope.l+JanSSTm.l+Depth.c+JanSSTm.l ²	5	-256.3375	522.7158	549.1536	0.0000	0.6168	0.0000	0.9554	42.9135	0.3569	0.9257	0.5330	0.8467	0.8741
PA ~ Slope.l+JanSSTm.l+Depth.c+JanSSTvar.c+JanSSTm.l ²	6	-255.6300	523.3172	555.0343	0.9521	0.3832	6.2314	0.0424	43.0711	0.3557	0.9254	0.5318	0.8474	0.8667
PA ~ Slope.l+Depth.c	3	-273.0408	552.0979	567.9688	29.5815	0.0000	18.4995	0.0001	39.1937	0.3247	0.9136	0.4972	0.8341	0.8741
PA ~ Slope.l+JanSSTvar.c	3	-333.1319	672.2801	688.1510	150.2661	0.0000	139.2166	0.0000	25.8114	0.1904	0.8584	0.3808	0.7670	0.8148
PA ~ Slope.l	2	-336.7513	677.5107	688.0940	155.1608	0.0000	138.9040	0.0000	25.0053	0.1842	0.8536	0.3778	0.7593	0.8000
PA ~ Depth.c	2	-353.3590	710.7261	721.3094	188.4702	0.0000	172.2746	0.0000	21.3068	0.1521	0.8314	0.3504	0.7022	0.8000
PA ~ JanSSTvar.c+JanSSTm.l+JanSSTm.l ²	4	-372.3683	752.7639	773.9196	230.2243	0.0000	224.6572	0.0000	17.0734	0.1116	0.8052	0.2957	0.7085	0.7407
MaxEnt (median)														
PA ~														
Slope.l+JanSSTm+Depth+JanSSTvar.c+Slope.l ² +JanSSTm ² +Depth ² +JanSSTvar.c ²	8	-714.7772	1445.9443	1485.3462	0.0000	0.6864	6.8351	0.0316	25.9188	0.3124	0.9209	0.5185	0.7919	0.9037
PA ~ Slope.l+JanSSTm+Depth+Slope.l ² +JanSSTm ² +Depth ²	6	-718.7651	1449.9570	1480.5641	1.5669	0.3136	0.0000	0.9611	25.5055	0.3132	0.9213	0.5156	0.7807	0.9185
PA ~ Slope.l+Depth+Depth ²	3	-738.8469	1485.8564	1506.2684	40.4540	0.0000	27.3045	0.0000	23.4242	0.2743	0.9125	0.4957	0.7630	0.9185
PA ~ Slope.l+JanSSTvar.c+Slope.l ² +JanSSTvar.c ²	4	-825.7029	1659.0461	1678.1177	214.0562	0.0000	199.4114	0.0000	14.4222	-0.0861	0.8564	0.3832	0.4681	0.9630
PA ~ Slope.l+Slope.l ²	2	-830.9622	1665.9493	1676.4112	221.0436	0.0000	197.2694	0.0000	13.8772	-0.1201	0.8525	0.3795	0.4570	0.9630
PA ~ Depth+Depth ²	2	-854.2826	1712.6496	1723.1434	266.8083	0.0000	243.7734	0.0000	11.4602	-0.3048	0.8306	0.3525	0.4222	0.9704
PA ~ JanSSTvar.c+JanSSTm+JanSSTvar.c ² +JanSSTm ²	4	-876.5274	1761.3649	1782.2378	316.2833	0.0000	302.9687	0.0000	9.1547	-0.6422	0.8075	0.2915	0.1237	1.0000
BRT (median)		deviance												
PA ~ Slope.l+JanSSTm.l+Depth.c+JanSSTm.l ²	-	526.6876	-	-	-	-	-	-	41.3532	0.3058	0.9094	0.4958	0.8644	0.7926
PA ~ Slope.l+JanSSTm.l+Depth.c+JanSSTvar.c+JanSSTm.l ²	-	526.4892	-	-	-	-	-	-	41.3753	0.3047	0.9091	0.4947	0.8644	0.7926
PA ~ Slope.l+Depth.c	-	542.9465	-	-	-	-	-	-	39.5428	0.2986	0.8994	0.4945	0.8622	0.7852
PA ~ Slope.l+JanSSTvar.c	-	644.0666	-	-	-	-	-	-	28.2830	0.1764	0.8223	0.3818	0.7889	0.7556
PA ~ Slope.l	-	657.9448	-	-	-	-	-	-	26.7377	0.1714	0.8129	0.3743	0.7859	0.7630
PA ~ Depth.c	-	695.0251	-	-	-	-	-	-	22.6088	0.1430	0.7986	0.3489	0.6907	0.7704
PA ~ JanSSTvar.c+JanSSTm.l+JanSSTm.l ²	-	708.2194	-	-	-	-	-	-	21.1396	0.0943	0.7763	0.2684	0.6844	0.7778

Table 2 5th percentiles of model metrics (Slope.l: log transformed bottom slope; JanSSTm.l: log-transformed January SST mean; JanSSTm.l²: the square of the log-transformed January SST mean; JanSSTvar.c: scaled interannual variance of January SST; Depth.c: scaled bathymetric depth; JanSSTm: January SST mean; Depth: bathymetric depth; LogL: log-likelihood; AICc: a sample size corrected AIC; dAICc: the difference between AIC value of a candidate model and the smallest AIC value in the set of candidate models; wAICc: AICc weight; BIC: Bayesian Information Criterion; dBIC: the difference between BIC value of a candidate model and the smallest BIC value in the set of candidate models; wBIC: BIC weight; BIC differences; pcdev: percentage of explained deviance; cvweight: CV weight; r2n_pred: cross-validated R²)

GLM (0.05)	k	LogL	AICc	BIC	dAICc	wAICc	dBIC	wBIC	pcdev	r2n_pred	AUC	Kappa	specificity	sensitivity
PA ~ Slope.l+JanSSTm.l+Depth.c+JanSSTm.l ²	5	-267.2309	496.6940	523.1318	0.0000	0.2235	0.0000	0.7942	40.4875	0.3281	0.9191	0.5006	0.8341	0.8593
PA ~ Slope.l+JanSSTm.l+Depth.c+JanSSTvar.c+JanSSTm.l ²	6	-266.3031	497.1890	528.9062	0.0000	0.2705	2.7890	0.0257	40.6942	0.3268	0.9188	0.5009	0.8348	0.8667
PA ~ Slope.l+Depth.c	3	-283.4283	527.8064	543.6773	20.2117	0.0000	9.1521	0.0000	36.8804	0.2974	0.9061	0.4669	0.8200	0.8667
PA ~ Slope.l+JanSSTvar.c	3	-342.3985	651.7846	667.6555	131.7768	0.0000	120.6196	0.0000	23.7477	0.1662	0.8494	0.3533	0.7526	0.8074
PA ~ Slope.l	2	-345.5559	658.1971	668.7805	136.8316	0.0000	120.6755	0.0000	23.0445	0.1621	0.8451	0.3483	0.7444	0.7852
PA ~ Depth.c	2	-360.1929	696.0172	706.6005	168.8261	0.0000	152.4504	0.0000	19.7849	0.1376	0.8221	0.3257	0.6859	0.8000
PA ~ JanSSTvar.c+JanSSTm.l+JanSSTm.l ²	4	-378.5760	738.4619	759.6176	209.2229	0.0000	203.0834	0.0000	15.6909	0.0982	0.7945	0.2707	0.6919	0.7259
MaxEnt (0.05)														
PA ~ Slope.l+JanSSTm+Depth+Slope.l ² +JanSSTm ² +Depth ²	6	-740.9223	1376.6198	1407.5116	0.0000	0.0001	0.0000	0.0053	23.2091	0.2625	0.9141	0.4847	0.7385	0.8741
PA ~														
Slope.l+JanSSTm+Depth+JanSSTvar.c+Slope.l ² +JanSSTm ² +Depth ² +JanSSTvar.c ²	8	-737.9638	1374.3864	1414.6660	0.0000	0.1113	0.0000	0.0007	23.5157	0.2615	0.9139	0.4865	0.7474	0.8593
PA ~ Slope.l+Depth+Depth ²	3	-758.3328	1430.3751	1450.8890	18.6652	0.0000	6.1037	0.0000	21.4046	0.2189	0.9052	0.4643	0.7199	0.8741
PA ~ Slope.l+JanSSTvar.c+Slope.l ² +JanSSTvar.c ²	4	-837.7823	1625.4989	1644.2664	172.0954	0.0000	156.9250	0.0000	13.1703	-0.1734	0.8478	0.3563	0.4370	0.9481
PA ~ Slope.l+Slope.l ²	2	-842.2029	1636.3092	1646.2314	178.2320	0.0000	155.4890	0.0000	12.7121	-0.2057	0.8439	0.3484	0.4281	0.9552
PA ~ Depth+Depth ²	2	-863.0877	1692.6760	1703.1619	223.1310	0.0000	199.9977	0.0000	10.5476	-0.3832	0.8210	0.3271	0.3785	0.9704
PA ~ JanSSTvar.c+JanSSTm+JanSSTvar.c ² +JanSSTm ²	4	-885.0697	1742.1623	1763.0351	275.6037	0.0000	261.3210	0.0000	8.2693	-0.7350	0.7959	0.2681	0.0696	1.0000
BRT (0.05)														
PA ~ Slope.l+JanSSTm.l+Depth.c+JanSSTm.l ²	-	506.6249	-	-	-	-	-	-	39.3689	0.2770	0.9010	0.4629	0.8466	0.7556
PA ~ Slope.l+JanSSTm.l+Depth.c+JanSSTvar.c+JanSSTm.l ²	-	506.3007	-	-	-	-	-	-	39.4005	0.2749	0.9007	0.4609	0.8459	0.7556
PA ~ Slope.l+Depth.c	-	521.1883	-	-	-	-	-	-	37.5516	0.2681	0.8902	0.4586	0.8385	0.7556
PA ~ Slope.l+JanSSTvar.c	-	626.2371	-	-	-	-	-	-	26.5830	0.1533	0.8105	0.3434	0.7555	0.7407
PA ~ Slope.l+dummy	-	640.3154	-	-	-	-	-	-	25.0535	0.1483	0.8018	0.3353	0.7429	0.7556
PA ~ Depth.c+dummy	-	679.6601	-	-	-	-	-	-	21.0525	0.1257	0.7858	0.3170	0.6429	0.7481
PA ~ JanSSTvar.c+JanSSTm.l+JanSSTm.l ²	-	694.2314	-	-	-	-	-	-	19.7063	0.0784	0.7609	0.2338	0.6570	0.7481

Table 3 95th percentiles of model metrics (Slope.l: log transformed bottom slope; JanSSTm.l: log-transformed January SST mean; JanSSTm.l²: the square of the log-transformed January SST mean; JanSSTvar.c: scaled interannual variance of January SST; Depth.c: scaled bathymetric depth; JanSSTm: January SST mean; Depth: bathymetric depth; LogL: log-likelihood; AICc: a sample size corrected AIC; dAICc: the difference between AIC value of a candidate model and the smallest AIC value in the set of candidate models; wAICc: AICc weight; BIC: Bayesian Information Criterion; dBIC: the difference between BIC value of a candidate model and the smallest BIC value in the set of candidate models; wBIC: BIC weight; BIC differences; pcdev: percentage of explained deviance; cvweight: CV weight; r2n_pred: cross-validated R^2)

	GLM (0.95)	k	LogL	AICc	BIC	dAICc	wAICc	dBIC	wBIC	pcdev	r2n_pred	AUC	Kappa	specificity	sensitivity
PA ~ Slope.l+JanSSTm.l+Depth.c+JanSSTvar.c+JanSSTm.l ²		6	-242.5659	544.6635	576.3807	1.9839	0.7765	7.2632	0.1979	45.9805	0.3918	0.9325	0.5621	0.8593	0.8741
PA ~ Slope.l+JanSSTm.l+Depth.c+JanSSTm.l ²		5	-243.3266	544.5027	570.9405	2.4903	0.7295	0.0000	0.9739	45.8111	0.3931	0.9327	0.5653	0.8585	0.8815
PA ~ Slope.l+Depth.c		3	-260.8950	572.8730	588.7439	41.4551	0.0000	30.3555	0.0094	41.8985	0.3568	0.9208	0.5299	0.8474	0.8889
PA ~ Slope.l+JanSSTvar.c		3	-322.8841	690.8132	706.6841	171.3755	0.0000	160.4095	0.0000	28.0935	0.2192	0.8674	0.4137	0.7800	0.8148
PA ~ Slope.l		2	-327.0945	695.1200	705.7033	176.1185	0.0000	159.8158	0.0000	27.1559	0.2111	0.8624	0.4044	0.7719	0.8074
PA ~ Depth.c		2	-346.0045	724.3940	734.9773	211.2550	0.0000	195.1356	0.0000	22.9446	0.1684	0.8407	0.3759	0.7178	0.8148
PA ~ JanSSTvar.c+JanSSTm.l+JanSSTm.l ²		4	-365.2173	765.1792	786.3349	253.3348	0.0000	247.9325	0.0000	18.6659	0.1269	0.8164	0.3247	0.7267	0.7556
MaxEnt (0.95)															
PA ~															
Slope.l+JanSSTm+Depth+JanSSTvar.c+Slope.l ² +JanSSTm ² +Depth ² +JanSSTvar.c ²		8	-678.7817	1492.2926	1531.8212	4.1550	0.9999	14.4748	0.9945	29.6495	0.3652	0.9280	0.5502	0.8437	0.9333
PA ~ Slope.l+JanSSTm+Depth+Slope.l ² +JanSSTm ² +Depth ²		6	-682.0068	1494.3604	1525.4178	18.8805	0.8886	10.4717	0.9992	29.3152	0.3677	0.9283	0.5488	0.8372	0.9407
PA ~ Slope.l+Depth+Depth ²		3	-711.1601	1524.7844	1545.4846	77.1610	0.0001	63.0024	0.0413	26.2937	0.3271	0.9200	0.5288	0.8104	0.9333
PA ~ Slope.l+JanSSTvar.c+Slope.l ² +JanSSTvar.c ²		4	-809.0662	1683.0365	1702.5169	274.9382	0.0000	261.7991	0.0000	16.1465	0.0157	0.8657	0.4190	0.5201	0.9704
PA ~ Slope.l+Slope.l ²		2	-816.1919	1688.4975	1698.9973	282.4769	0.0000	259.9702	0.0000	15.4080	-0.0222	0.8616	0.4057	0.5089	0.9633
PA ~ Depth+Depth ²		2	-844.2922	1730.2611	1740.7442	337.1888	0.0000	314.5429	0.0000	12.4956	-0.2260	0.8398	0.3769	0.4681	0.9704
PA ~ JanSSTvar.c+JanSSTm+JanSSTvar.c ² +JanSSTm ²		4	-866.9261	1778.4496	1799.3224	387.1690	0.0000	374.6654	0.0000	10.1498	-0.5462	0.8187	0.3194	0.1719	1.0000
BRT (0.95)			deviance												
PA ~ Slope.l+JanSSTm.l+Depth.c+JanSSTm.l ²		-	544.5077	-	-	-	-	-	-	43.5872	0.3375	0.9176	0.5340	0.8844	0.8222
PA ~ Slope.l+JanSSTm.l+Depth.c+JanSSTvar.c+JanSSTm.l ²		-	544.2239	-	-	-	-	-	-	43.6233	0.3362	0.9175	0.5321	0.8837	0.8222
PA ~ Slope.l+Depth.c		-	560.8286	-	-	-	-	-	-	41.9656	0.3309	0.9084	0.5338	0.8837	0.8222
PA ~ Slope.l+JanSSTvar.c		-	659.3344	-	-	-	-	-	-	30.2683	0.2065	0.8345	0.4224	0.8119	0.7778
PA ~ Slope.l		-	673.0703	-	-	-	-	-	-	28.7007	0.2025	0.8236	0.4121	0.8134	0.7781
PA ~ Depth.c		-	709.0020	-	-	-	-	-	-	24.3197	0.1648	0.8108	0.3769	0.7615	0.8667
PA ~ JanSSTvar.c+JanSSTm.l+JanSSTm.l ²		-	721.0912	-	-	-	-	-	-	22.6972	0.1144	0.7917	0.3193	0.7193	0.8074