

曹睿星, 官文江, 高峰, 等. 基于最大熵和栖息地指数模型预测东、黄海日本鲭渔场分布[J]. 海洋学报, 2023, 45(9): 72–81, doi:10.12284/hyxb2023136

Cao Ruixing, Guan Wenjiang, Gao Feng, et al. Prediction of chub mackerel fishing ground distribution in the East China Sea and Yellow Sea based on maximum entropy model and habitat suitability index model[J]. Haiyang Xuebao, 2023, 45(9): 72–81, doi:10.12284/hyxb2023136

基于最大熵和栖息地指数模型预测东、 黄海日本鲭渔场分布

曹睿星¹, 官文江^{1,2}, 高峰^{1,3*}, 贺伟伟¹

(1. 上海海洋大学 海洋科学学院, 上海 201306; 2. 大洋渔业资源可持续开发教育部重点实验室, 上海 201306; 3. 上海海洋大学 国家远洋渔业工程技术研究中心, 上海 201306)

摘要: 最大熵模型 (Maximum Entropy Model, Maxent) 和栖息地指数 (Habitat Suitability Index, HSI) 模型均广泛应用于渔情预报研究中。为比较两模型渔情预报效果以提升日本鲭 (*Scomber japonicus*) 资源的科学管理水平, 本研究利用 2003–2012 年东、黄海日本鲭的渔业数据以及海表温度、海面高度、海表盐度、海表温度梯度等海洋环境数据, 构建最大熵模型和 HSI 模型, 以分析、比较两模型对东、黄海日本鲭栖息地的预测效果, 并利用受试者工作特征 (Receiver Operating Characteristic, ROC) 曲线下面积 (Area Under Curve, AUC)、模型预测的渔场概率与实际渔获量百分比之间的对应关系对两模型渔情预报效果进行了定量评价。结果表明: (1) 最大熵模型预测的渔场发生高概率位置与捕捞位置基本重合, 在无历史捕捞数据海域预测渔场发生的概率较低; HSI 模型预测的高栖息地指数位置与捕捞位置部分重合, 在无历史捕捞数据海域也可获得较高的栖息地指数, 将非渔场预测为渔场的概率较高; (2) 最大熵模型和 HSI 模型的月平均 AUC 值分别为 0.95 和 0.66, 故最大熵模型的预测结果相对较好; (3) 使用 HSI 模型时, 应在模型中加入非渔场数据, 并加强对此类数据的收集, 否则该类模型预报渔场时有扩大化的可能; 使用最大熵模型时, 必须提高渔业数据的空间覆盖率, 否则无法全面反映渔场时空分布动态。本文研究结果可为提升东、黄海日本鲭渔情预报精度提供参考。

关键词: 最大熵模型; 栖息地指数模型; 模型比较; 日本鲭

中图分类号: P714+.5; S932.4

文献标志码: A

文章编号: 0253-4193(2023)09-0072-10

1 引言

日本鲭 (*Scomber japonicus*) 广泛分布于大西洋、太平洋、印度洋等沿岸海域^[1], 分布于东、黄海的日本鲭是我国重要的中上层经济鱼类^[2]。东、黄海日本鲭的资源丰度和空间分布受海洋环境影响显著, 其栖息地和资源量的年际变化较为明显^[3]。因此, 利用海洋环境因子构建较为精确的渔情预报模型, 对可持续利

用和科学管理日本鲭资源有积极作用。

目前, 用于渔情预报的渔业数据主要分为两类, 一类是“存在-缺失”(presence-absence) 数据, 该类数据假设非渔场位置也有足够观测样本; 另一类是“仅存在”(presence-only) 数据, 即“渔场”有观测数据, 而“非渔场”则无观测数据^[4]。由于商业捕捞的作业位置受经验、探鱼技术等多种因素影响而不具备随机性, 因此大部分渔业数据具有“仅存在”数据性质^[5], 目前

收稿日期: 2023-03-22; 修订日期: 2023-06-26。

基金项目: 国家自然科学基金 (32072981)。

作者简介: 曹睿星 (1999—), 女, 安徽省无为市人, 研究方向为渔业资源评估与管理。E-mail: 13955930469@163.com

* 通信作者: 高峰, 男, 讲师, 研究方向为渔业地理信息系统与渔业海图。E-mail: gaofeng@shou.edu.cn

已有的日本鲭渔业数据也是“仅存在”数据^[6]。故用基于“存在-缺失”数据的模型对渔场进行预测时, 预测结果可能会存在较大误差^[2]。

国内外学者对日本鲭渔情预报的研究较多, 目前常用的日本鲭渔情预报模型包括广义加性模型 (Generalized Additive Models, GAM)^[7]、提升回归树 (Boosted Regression Trees, BRT) 模型^[2]、栖息地指数 (Habitat Suitability Index, HSI) 模型^[8-9] 以及最大熵模型 (Maximum Entropy Model, Maxent)^[10-12] 等, 但这些研究尚未考虑渔业数据的特殊性对两类模型预测效果的影响。目前, 在可用于“仅存在”数据的渔情预报模型中, 较为常见的是栖息地指数模型和最大熵模型。两者不同之处在于最大熵模型是专门为“仅存在”数据而开发和建立的模型, 不需要“不存在”数据的明确约束^[13-15]。栖息地指数模型则最早应用于“存在-缺失”数据^[16], 而当应用于“仅存在”数据时, 容易产生过度预测倾向^[17]。

本研究根据 2003–2012 年东、黄海日本鲭渔业数据及相关海洋环境因子, 建立基于最大熵和栖息地指数的日本鲭渔情预报模型, 通过分析两模型对日本鲭栖息地预测的结果与受试者工作特征 (Receiver Operating Characteristic, ROC) 曲线下的面积 (Area Under Curve, AUC), 及两模型预测的渔场概率与实际渔获量百分比之间的对应关系, 比较两类模型在渔情预报性能上的差异。本研究将为日本鲭渔情预报模型的完善, 提高东、黄海日本鲭渔情预报精度提供参考。

2 材料与方法

2.1 数据来源

中国大型机轮灯光围网渔业的日本鲭捕捞数据来自上海海洋大学鱿钩技术组, 时间为 2003–2012 年, 区域为 25°~40°N、120°~130°E 范围内的东、黄海海域。渔业数据包括生产日期、船名、作业位置、捕捞网次、捕捞产量等, 数据的时间分辨率为月, 空间分辨率为 (1/6)°。海洋环境数据包括海表温度 (Sea Surface Temperature, SST)、海面高度 (Sea Surface Height, SSH)、海表盐度 (Sea Surface Salinity, SSS) 等, 这些数据均来自于哥白尼海洋环境监测服务中心 (Copernicus-Marine Environment Monitoring Service, CMEMS, <https://resources.marine.copernicus.eu>), 数据的时间分辨率均为月, 空间分辨率为 0.25°。

海表温度梯度 (Sea Surface Temperature Gradient, STG) 的计算公式为

$$STG_{i,j} = \sqrt{\left[\frac{(\text{SST}_{i,j-1} - \text{SST}_{i,j+1})^2 + (\text{SST}_{i+1,j} - \text{SST}_{i-1,j})^2}{2} \right]}, \quad (1)$$

式中, i 和 j 分别表示 SST 在 0.25° × 0.25° 网格的行、列号。

将海洋环境数据插值到 (1/6)° × (1/6)° 的网格内, 并使之与渔业数据相匹配, 以用于模型的构建与日本鲭栖息地的预测。本研究选取 2003–2011 年的渔业数据作为训练数据集以建立预报模型, 选取 2012 年的渔业数据作为测试数据集以比较预测效果。

2.2 模型构建

2.2.1 最大熵模型

最大熵模型是以最大熵理论为基础而建立的, 最大熵理论最早由 Jaynes^[18] 提出, 其中熵是对整个范围内随机变量不确定性的量度^[19], 熵最大时, 说明随机变量最不确定。最大熵理论的实质是利用已知数据有限信息推断未知的概率分布, 关于未知分布最合理的推断是使其符合已知信息最随机的推断 (即假设未知信息呈均匀分布), 即熵最大化。任何其他的推断都将对未知分布增加约束条件与假设, 故只有当此概率分布最均匀、熵最大 (即符合上述最合理的推断) 时, 模型误差最低^[20]。因此, 通过熵的最大化来实现未知信息最均匀的分布^[19], 可最大限度排除主观因素干扰, 从而达到估计物种分布概率的目的^[21]。

假设一个有限集合 X 上的未知概率分布为 $P(X)$, 其中 $x_1, x_2, x_3, \dots, x_m$ (m 表示训练点个数) 表示在集合 X 上的训练样本, 由集合 X 上的一组特征 $f_1, f_2, f_3, \dots, f_n$ (n 为环境因子个数) 表示未知概率分布 $P(X)$ 的约束条件, 这些特征是对有限集合 X 的不完整描述^[22]。为了估计未知概率分布 $P(X)$, 需要构建一个近似于 $P(X)$ 的经验分布 $\tilde{P}(X)$, $\tilde{P}(X)$ 可以表述为

$$\tilde{P}(X) = \frac{|\{1 \leq i \leq m : x_i = x\}|}{m}, \quad (2)$$

式中, x_i 表示在集合 X 上的训练样本; m 为训练点个数。

特征函数 f_j 关于经验分布 $\tilde{P}(X)$ 的期望, 用 $E_{\tilde{P}}(f_j)$ 表示:

$$E_{\tilde{P}}(f_j) = \frac{\sum_{i=1}^m f_j(x_i)}{m}, \quad (3)$$

假设 $P(X)$ 的期望与 $\tilde{P}(X)$ 的期望相同, 则 $P(X)$ 的熵 $H(P)$ 定义为式 (4)。最大熵模型就是在所有满足已知约束条件的模型集合中选取熵最大的模型^[22], 其估计的物种分布概率由式 (5) 计算。

$$H(P) = - \sum_{x \in X} P(x) \ln P(x), \quad (4)$$

$$P(x) = \frac{\sum_{j=1}^n \lambda_j f_j(x)}{Z}, \quad (5)$$

式中, λ_j 为不同环境因子权重, 可通过式(3)与式(5)计算; $f_j(x)$ 为第 j 个环境因子的特征函数; Z 为归一化常数。

本文在构建最大熵模型时, 首先利用渔业捕捞数据构建样本点坐标数据集; 其次, 对多年环境数据按月进行平均, 生成环境图层数据; 运用 MaxEnt 软件(3.4.1)^[23] 建立最大熵模型, 并利用 ArcMap 软件(10.6.1) (<http://desktop.arcgis.com/zh-cn/>) 生成渔业资源的分布概率图^[24-27]。

2.2.2 HSI 模型

本文假设作业次数与日本鲭资源量呈正比关系^[28], 适应性指数(Suitability Index, SI)计算公式为

$$SI_i = \frac{FT_i}{FT_{i,max}}, \quad (6)$$

式中, FT_i 为 i 月各捕捞点的作业次数; $FT_{i,max}$ 为 i 月的最大作业次数。

SI 与环境因子的关系如下^[29]:

$$SI_k = e^{-a \ln(x_k - b)^2}, \quad (7)$$

式中, a 和 b 为待估参数; x_k 为环境因素 k , 即 SST、SSH、SSS、STG 的值。

通过上述方程的拟合, 得到每月各环境因子的适宜性曲线方程, 将每月数据代入相应方程, 得到不同月份环境因子所对应的 SI。本研究先、后利用算术平均法(Arithmetic Mean Model, AMM)和几何平均法(Geometric Mean Model, GMM)计算栖息地指数:

$$HSI = \frac{(SI_{SST} + SI_{SSH} + SI_{SSS} + SI_{STG})}{4}, \quad (8)$$

$$HSI = \sqrt[4]{SI_{SST} \times SI_{SSH} \times SI_{SSS} \times SI_{STG}}, \quad (9)$$

式中, SI_{SST} 、 SI_{SSH} 、 SI_{SSS} 和 SI_{STG} 分别为 SST、SSH、SSS 和 STG 的适宜性指数。

对比 AMM 与 GMM 在 7-12 月对应的产量分布以选取合适的算法进行 HSI 模型的计算, 将东、黄海日本鲭的渔业数据与计算得到的 HSI 值进行叠加, 绘制东、黄海日本鲭栖息地分布图。

2.3 模型比较

本研究利用 ROC 曲线以及曲线下面积(即 AUC)对模型的渔情预报效果进行评价。其中, ROC 曲线的横坐标为假阳性率(即 1-特异性), 即实际渔场不存在, 但模型预测为存在的样本(即反例样本)在总反例样本中的占比; 纵坐标为真阳性率(即灵敏度), 即模型正确预测渔场存在的样本(即正例样本)在总正例

样本中的占比^[30-31]。ROC 曲线越接近顶点(0, 1), 表明环境因子与预测的渔场分布模型之间相关性越大, 即模型预报精度越高^[31-33]。AUC 值的范围在 0.5~1 之间, 当 AUC 值为 0.5~0.6 时模型预测精度较差; 当 AUC 值为 0.6~0.7 时, 表明模型预测精度一般; 当 AUC 值为 0.7~0.8 时模型预测精度较准确; 当 AUC 值为 0.8~0.9 时, 表明模型精度很准确; 当 AUC 值大于 0.9 时, 表明预测结果极准确^[27]。

最大熵模型的 ROC 曲线和 AUC 值可从 MaxEnt 软件输出的结果中直接获取^[34]; MaxEnt 软件在计算 AUC 值时, 由于没有真实“不存在”数据, 故需从研究区域中均匀随机选择背景数据代之进行计算^[26]; 对于栖息地指数模型需先利用分位数法对各月捕捞努力量进行二值化处理, 即需要选择一个阈值将渔区分为低值区和高值区两类, 并用 0(低值区)与 1(高值区)表示^[35-36]。阈值选择依据是将逻辑值为 1 的点叠加到不同阈值的栖息地指数分布图上, 判别高栖息地指数位置与高值渔区的分布是否吻合, 从而选出最优阈值, 本研究最终以 50% 分位数作为分割点对各月捕捞努力量进行分级。ROC 曲线的绘制和 AUC 值的相关计算使用 R 语言^[37] 中的 Presence Absence 包^[38] 实现。同时, 统计分析两类模型预测的渔场概率与实际渔获量百分比之间的对应关系。

3 结果

3.1 栖息地指数模型选择

在栖息地指数的计算中, 本文先、后对比了 AMM 和 GMM 在 7-12 月所对应的渔获量分布。结果显示(表 1), 对于 AMM, 39.13% 的渔获量集中在 $HSI < 0.4$ 的区间, 33.61% 的渔获量集中在 $HSI \geq 0.6$ 的区间; 对于 GMM, 65.95% 的渔获量集中在 $HSI < 0.4$ 的区间, 20.13% 的渔获量集中在 $HSI \geq 0.6$ 的区间。由此可见, 在本研究中, AMM 的预测效果更好, 故最终选取 AMM 计算栖息地指数。

3.2 渔场预测

如图 1 所示, 最大熵模型预测(即用 2012 年数据所做的预测)的高概率渔场在 7 月主要分布于东海中南部及舟山外海; 自 9 月开始, 高概率渔场逐步由东海南部渔场转移至东海北部及黄海中部渔场; 10 月, 高概率渔场主要集中在黄海中北部, 并在 11 月与 12 月, 逐步向黄海中南部转移; 12 月份高概率渔场在东海北部与东海南部有所增加。与实际捕捞位置相比, 除 8 月及 9 月少数捕捞位置不在预测的较高渔场发生概率区外, 其他捕捞位置与最大熵模型预测的高

表 1 利用 AMM 和 GMM 分别构建栖息地指数模型的作业次数比重的比较

Table 1 Comparison of the weight of the number of operations for constructing habitat index model using AMM and GMM respectively

HSI	7月作业次数 比重/%		8月作业次数 比重/%		9月作业次数 比重/%		10月作业次数 比重/%		11月作业次数 比重/%		12月作业次数 比重/%		平均值作业 次数比重/%	
	AMM	GMM	AMM	GMM	AMM	GMM	AMM	GMM	AMM	GMM	AMM	GMM	AMM	GMM
[0, 0.2)	3.37	49.44	0.00	4.19	39.88	85.28	10.53	23.16	27.27	87.02	2.99	62.69	14.01	51.96
[0.2, 0.4)	10.11	21.35	4.79	23.95	48.46	13.50	8.42	7.37	46.10	5.84	32.83	11.94	25.12	13.99
[0.4, 0.6)	41.57	6.75	27.55	25.15	11.66	1.22	35.79	35.79	20.14	7.14	26.86	7.46	27.26	13.92
[0.6, 0.8)	35.96	13.48	43.11	30.54	0.00	0.00	34.74	26.31	6.49	0.00	32.84	17.91	25.52	14.71
[0.8, 1.0]	8.99	8.99	24.55	16.17	0.00	0.00	10.52	7.37	0.00	0.00	4.48	0.00	8.09	5.42

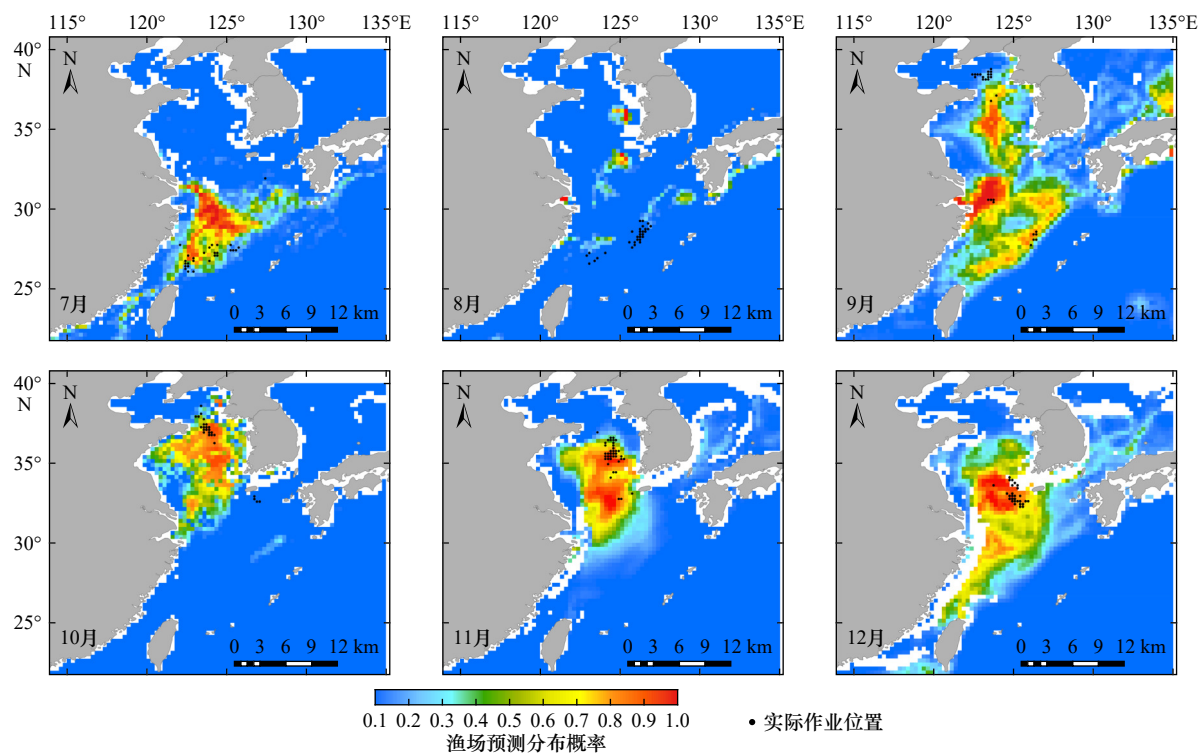


图 1 基于 Maxent 预测的渔场分布概率与捕捞作业位置

Fig. 1 Distribution probability of fishing grounds and location of fishing operation predicted by maximum entropy model

概率渔场位置有较好的重合。对无历史作业位置的海域(如日本海、渤海等海域),最大熵模型预测的渔场发生概率较低(图 1),如除 9 月外,其他月份在日本海、渤海海域的渔场发生概率都很低。

栖息地指数模型预测的高概率渔场在 7 月、8 月、9 月主要分布于东海南部及日本海(图 2);自 10 月开始,高概率渔场逐步由东海南部渔场转移至渤海中南部及黄海中部渔场;并在 11 月与 12 月,逐步向黄海西南部及日本海转移。栖息地指数模型预测的高栖息地指数位置与捕捞位置部分重合(图 2),如在 9 月、10 月及 12 月模型预测的高栖息地指数位

置在部分海域(如 9 月的黄海中部,10 月的东海海域)与实际捕捞作业分布存在较大差异。与最大熵模型相比,栖息地指数模型预测的渔场分布范围(高栖息地指数海域)明显较大(图 2),在无历史作业位置的海域(如日本海、渤海等海域),其栖息地指数也较高,如在 7 月、8 月、9 月的日本海海域,以及 10 月的渤海海域,其栖息地指数的预测值均较高。

3.3 模型精度

最大熵模型的 ROC 曲线更接近顶点(0, 1),远离随机分布预测模型,说明该模型预测的准确率更高(图 3)。相比之下,栖息地指数模型的 ROC 曲线与随

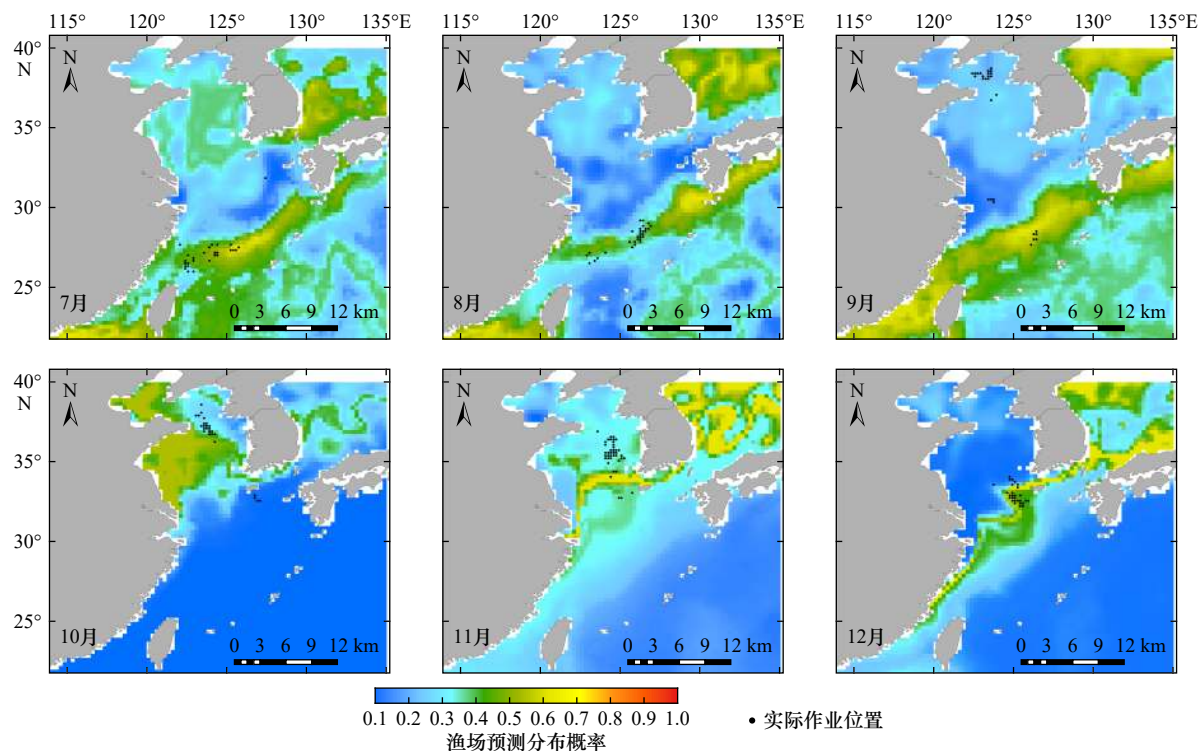


图 2 HSI 模型预测的渔场分布概率与捕捞作业位置

Fig. 2 Distribution of habitat suitability index and locations of fishing operation predicted by the Habitat Suitability Index model

机分布预测模型的距离较近,且在相同的真阳性率下,该模型的假阳性率更高,表明该模型在渔场预测中将更多实际不存在的渔场预测为渔场(图 3)。

如表 2 所示,利用最大熵模型得到的每个月 AUC 值均大于 0.9,即模型预测精度相对较高;利用栖息地指数模型得到的每个月的 AUC 值在 0.5~0.8

之间,即模型预测精度一般。最大熵模型和栖息地指数模型的月平均 AUC 值分别为 0.95 和 0.66(表 2),最大熵模型的 AUC 值远高于栖息地指数模型,最大熵模型的预测效果相对较好。

对于最大熵模型,大部分作业位于预测渔场概率 0.4~0.8 的区域(图 4),随着预测渔场概率的增大,渔

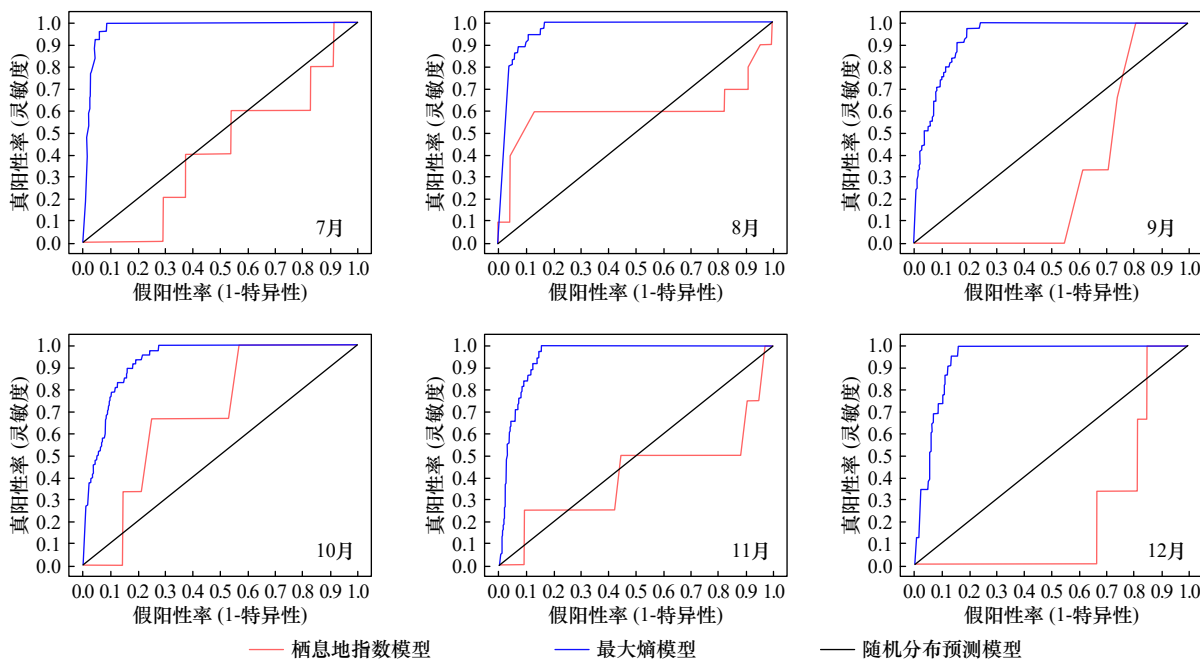


图 3 栖息地指数模型和最大熵模型的 ROC 曲线

Fig. 3 ROC curve of habitat suitability index model and maximum entropy model

表 2 基于两种模型不同月份的 AUC 值

Table 2 The AUC values on different months based on two models

模型	7月	8月	9月	10月	11月	12月	平均
最大熵模型	0.98	0.96	0.93	0.93	0.95	0.94	0.95
栖息地指数模型	0.59	0.60	0.68	0.69	0.60	0.78	0.66

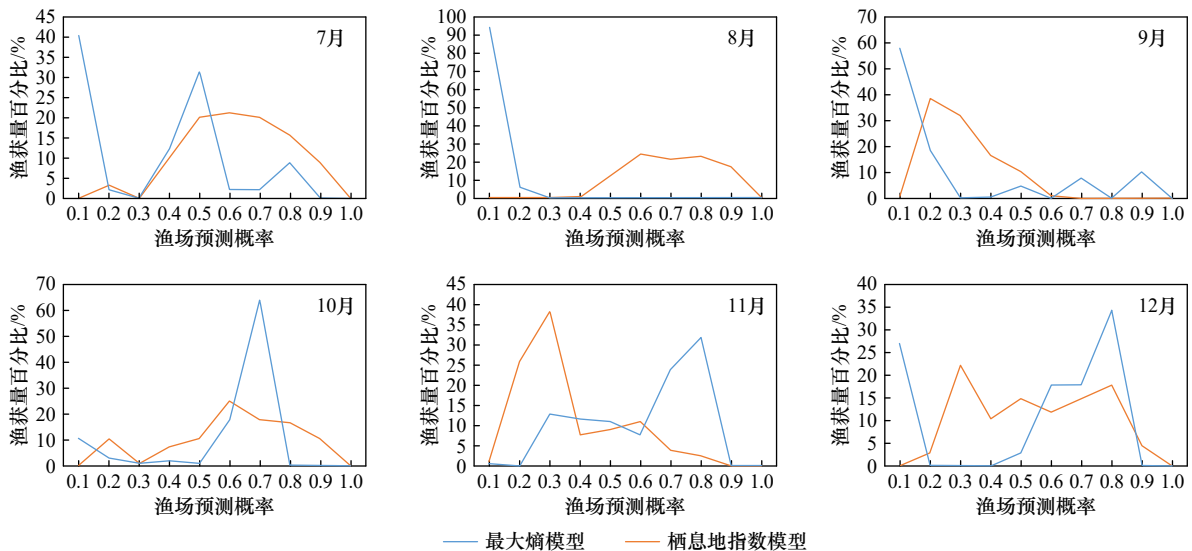


图 4 实际作业渔获量百分比与两个模型预测渔场概率的关系

Fig. 4 Relationship between the percentage of total catch and the predicted probability of fishing grounds by two models

获量百分比先增加后减少,最大渔获量百分比大多位于预测渔场概率 0.6~0.8 的区域。而对栖息地指数模型而言,大部分作业位于预测渔场概率 0.2~0.6 的区域(图 4),随着预测渔场概率的增大,渔获量百分比先增加后减少,7月、8月、10月最大渔获量百分比位于预测渔场概率 0.5~0.7 的区域,9月、11月、12月最大渔获量百分比位于预测渔场概率 0.2~0.3 的区域。10-12月,最大熵模型预测的渔场概率集中在大于 0.6 的区域(图 4),明显强于栖息地指数模型;但在 7-9月,栖息地指数模型略强于最大熵模型(图 4)。

4 讨论

4.1 模型对比分析

(1) 模型原理

最大熵模型以最大熵理论为基础,利用“仅存在”的物种分布与环境数据预测物种潜在分布范围,计算物种分布点的环境因子对该物种构成的约束条件,并寻找在约束条件下熵最大的物种分布概率^[39-40]。同时,最大熵模型利用所有已知数据进行建模,对未知数据作出呈“均匀分布”的假设,并通过熵的最大化来表示此类均匀分布^[41-42]。栖息地指数模型最早应用于

“存在-缺失”数据^[43],但也用于“仅存在”数据^[44],此时,由于没有“不存在”数据约束模型,HSI模型不可避免倾向于过度预测适宜栖息地^[17]。相较而言,栖息地指数模型预测的渔场范围相对较大而分散,故实际捕捞作业点处在最适宜栖息地范围内的概率也相对较高(图 4),导致该模型预测的高渔场概率位置处的渔获量百分比偏高。

(2) 渔业数据特性

由于我国大型机轮灯光围网渔业长期在东、黄海捕捞日本鲭,捕捞从业者对东、黄海日本鲭的时空分布动态相对较为了解,其捕捞位置通常位于日本鲭渔场。因此,我国大型机轮灯光围网渔业的数据主要表现为“仅存在”数据。但由于东、黄海日本鲭被我国大陆及台湾省、韩国、日本等共同捕捞,我国大型机轮灯光围网渔业数据并不能完全覆盖东、黄海日本鲭渔场位置,故实际上最大熵模型预测的结果不能完全反映东、黄海日本鲭渔场的空间分布,如 11 月份,在我国浙江近海有大量群众渔业捕捞日本鲭,该海域也是重要的日本鲭渔场,但图 1 无法反映这一点。同时,由于 7-9 月在东海中南部捕捞作业较少,导致最大熵模型预测的最适宜栖息地范围相对较小,因此,实际捕捞作业位置在最适预测范围内的概率也相对

较小, 最终导致渔获量百分比偏小(图 4)。

根据以上两方面的分析可知, 使用栖息地指数模型时, 应考虑将非渔场数据加入到 HSI 模型中, 这需要加强对非渔场数据的收集, 即也需要对无捕捞海域进行采样, 否则必然使该类模型预报的渔场有较大可能扩大化, 导致第二类错误的产生, 即非渔场也可能被预测为渔场。但考虑到经济成本等多方面因素的影响, 此方法较难实现^[4]。目前, 渔业数据多为“仅存在”数据, 由国内外研究现状可知, 基于该类数据的模型较少, 而在仅有的这些模型中, 最大熵模型显示出比其他模型更高的预测精度^[45]。模型检验的结果(表 2)也表明最大熵模型对实际渔场拟合与预报效果相较于栖息地指数模型表现更好。故在今后渔情预报的研究中, 应加强发展与使用基于“仅存在”数据的模型, 如加强使用最大熵模型。同时, 应注意到, 使用最大熵模型时必须提高渔业数据的空间覆盖率, 否则也将不能全面反映渔场的时空分布动态^[46]。

4.2 环境数据的选择

影响日本鲭时空分布的环境因子较多, 若选用了日本鲭空间分布影响不显著的环境因子或缺少具有重要影响的环境因子, 均可能影响模型预测效果^[47]。除本文使用的环境因子外, 还有如叶绿素 *a* 浓度、海洋初级生产力等数据能反映日本鲭的饵料分布, 因此, 这些因子对日本鲭的空间分布可能有重要影响, 加入这些因子有利于提高日本鲭空间分布的预测。但由于东、黄海海域属于二类水体, 基于遥感的叶绿素 *a* 浓度、海洋初级生产力等产品存在较大误差, 这有可能影响模型的预测精度^[2]。鉴于本文旨在比较两类模型对渔场预测的影响及该类数据存在较大不确定性等原因, 本文没有使用这些数据。

4.3 适应性指数模型的构建

一般, 假设单位捕捞努力渔获量 (Catch Per Unit

Effort, CPUE) 与渔业资源量呈正比关系, 因此, 在构建适应性指数模型时, 通常采用归一化 CPUE 作为 SI^[48-49]。但由于我国大型机轮灯光围网渔业长期在东、黄海捕捞日本鲭, 捕捞从业者对日本鲭渔场分布及其时空动态有较多了解, 在此条件下, 根据理想自由分布理论^[50], 日本鲭的 CPUE 存在被均匀化的可能, CPUE 不能较好反映日本鲭资源量的变化, 而捕捞努力量与资源量则呈正比关系^[19]。据此, 本文在构建 SI 时采用了捕捞努力量数据。

4.4 二值化处理的阈值选择

本研究在利用分位数法对各月捕捞努力量进行分级处理时, 分别选取了 25%、50% 和 75% 分位数作为分割点, 进行高值渔区和低值渔区的区分。将逻辑值为 1 的点叠加到栖息地指数分布图上, 判别高栖息地指数位置与高值渔区的分布是否吻合。结果显示, 当以 50% 作为分割点对捕捞努力量进行分级时, 得到的结果与实际情况最为吻合, 故本研究最终确定以 50% 分位数作为分割点进行二值化处理。

5 结语

本研究利用中国大型机轮灯光围网渔业的日本鲭捕捞数据, 建立了基于最大熵和栖息地指数的日本鲭渔场预报模型, 并对东、黄海日本鲭栖息地适宜性分布进行了预测, 比较了两模型在适宜性栖息地预测中的差异。由于最大熵模型和栖息地指数模型基于不同的模型假设及数据需求, 两模型的预测效果差异明显, 对本研究所使用的“仅存在”渔业数据而言, 最大熵模型的预测效果更好。在后续渔情预报研究中, 应结合渔业数据的特点及模型本身的假设与数据需求建立渔情预报模型, 并结合多模型的预测优势建立综合渔情预报模型, 以提升东、黄海日本鲭渔情预报精度。

参考文献:

- [1] 官文江, 马雪莲. 利用贝叶斯动态产量模型评估东、黄海日本鲭资源状况[J]. 上海海洋大学学报, 2022, 31(3): 749-760.
Guan Wenjiang, Ma Xuelian. Assessment of the status of *Scomber japonicus* resources in the East China Sea and Yellow Sea using a Bayesian biomass dynamic model[J]. Journal of Shanghai Ocean University, 2022, 31(3): 749-760.
- [2] 高峰. 基于提升回归树的东、黄海鲈鱼渔场预报模型研究[D]. 上海: 上海海洋大学, 2016.
Gao Feng. Fishing ground forecasting of chub mackerel in the East China Sea and Yellow Sea using boosted regression trees[D]. Shanghai: Shanghai Ocean University, 2016.
- [3] 郭爱. 气候与海洋环境变化对东黄海鲈鱼栖息地时空变动的影响[D]. 上海: 上海海洋大学, 2020.
Guo Ai. Impacts of the climatic and environmental variations on the spatio-temporal distribution of potential habitat of chub mackerel *Scomber japonicus* in the East China Sea and Yellow Sea[D]. Shanghai: Shanghai Ocean University, 2020.
- [4] Phillips S J, Anderson R P, Schapire R E. Maximum entropy modeling of species geographic distributions[J]. Ecological Modelling, 2006, 190(3/4): 231-259.
- [5] 陈新军, 高峰, 官文江, 等. 渔情预报技术及模型研究进展[J]. 水产学报, 2013, 37(8): 1270-1280.

- Chen Xinjun, Gao Feng, Guan Wenjiang, et al. Review of fishery forecasting technology and its models[J]. *Journal of Fisheries of China*, 2013, 37(8): 1270–1280.
- [6] Wang Lifei, Kerr L A, Record N R, et al. Modeling marine pelagic fish species spatiotemporal distributions utilizing a maximum entropy approach[J]. *Fisheries Oceanography*, 2018, 27(6): 571–586.
- [7] 杨胜龙, 范秀梅, 伍玉梅, 等. 基于GAM模型的阿拉伯海鲈鱼渔场分布与环境关系[J]. *生态学杂志*, 2019, 38(8): 2466–2470.
Yang Shenglong, Fan Xiumei, Wu Yumei, et al. The relationship between the fishing ground of mackerel (*Scomber australasicus*) in Arabian Sea and the environment based on GAM model[J]. *Chinese Journal of Ecology*, 2019, 38(8): 2466–2470.
- [8] 易炜, 郭爱, 陈新军. 不同环境因子权重对东海鲈鱼栖息地模型的影响研究[J]. *海洋学报*, 2017, 39(12): 90–97.
Yi Wei, Guo Ai, Chen Xinjun. A study on influence of different environmental factors weights on the habitat model for *Scomber japonicus*[J]. *Haiyang Xuebao*, 2017, 39(12): 90–97.
- [9] Lee D, Son S, Kim W, et al. Spatio-temporal variability of the habitat suitability index for chub mackerel (*Scomber Japonicus*) in the East/Japan Sea and the South Sea of South Korea[J]. *Remote Sensing*, 2018, 10(6): 938.
- [10] 刘思源, 张衡, 杨超, 等. 基于最大熵模型的西北太平洋远东拟沙丁鱼和日本鲭栖息地差异[J/OL]. [2023–05–05]. 上海海洋大学学报, 2023: 1–13. <http://kns.cnki.net/kcms/detail/31.2024.S.20230421.1814.004.html>.
Liu Siyuan, Zhang Heng, Yang Chao, et al. Differences in habitat distribution of *Sardinops melanostictus* and *Scomber japonicus* in the Northwest Pacific based on a maximum entropy model[J/OL]. [2023–05–05]. *Journal of Shanghai Ocean University*, 2023: 1–13. <http://kns.cnki.net/kcms/detail/31.2024.S.20230421.1814.004.html>.
- [11] 陈芑, 陈新军. 基于最大熵模型分析西南大西洋阿根廷滑柔鱼栖息地分布[J]. *水产学报*, 2016, 40(6): 893–902.
Chen Peng, Chen Xinjun. Analysis of habitat distribution of Argentine shortfin squid (*Illex argentinus*) in the southwest Atlantic Ocean using maximum entropy model[J]. *Journal of Fisheries of China*, 2016, 40(6): 893–902.
- [12] 张嘉容, 杨晓明, 田思泉. 基于最大熵模型的南太平洋长鳍金枪鱼栖息地预测[J]. *中国水产科学*, 2020, 27(10): 1222–1233.
Zhang Jiarong, Yang Xiaoming, Tian Siqun. Analysis of albacore (*Thunnus alalunga*) habitat distribution in the South Pacific using maximum entropy model[J]. *Journal of Fishery Sciences of China*, 2020, 27(10): 1222–1233.
- [13] Monk J, Ierodionou D, Versace V L, et al. Habitat suitability for marine fishes using presence-only modelling and multibeam sonar[J]. *Marine Ecology Progress Series*, 2010, 420: 157–174.
- [14] Fourcade Y, Engler J O, Rödder D, et al. Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias[J]. *PLoS One*, 2014, 9(5): e97122.
- [15] 龚彩霞, 陈新军, 高峰. 基于最大熵模型模拟西北太平洋柔鱼潜在栖息地分布[J]. *中国水产科学*, 2020, 27(3): 336–345.
Gong Caixia, Chen Xinjun, Gao Feng. Modeling the potential distribution of the neon flying squid (*Ommastrephes bartramii*) in the Northwest Pacific Ocean based on a MaxEnt model[J]. *Journal of Fishery Sciences of China*, 2020, 27(3): 336–345.
- [16] Lee P Y, Suen J P. Comparing habitat suitability indices (HSIs) based on abundance and occurrence data[J]. *North American Journal of Fisheries Management*, 2013, 33(1): 89–96.
- [17] Zaniwski A E, Lehmann A, Overton J M. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns[J]. *Ecological Modelling*, 2002, 157(2/3): 261–280.
- [18] Jaynes E T. Information theory and statistical mechanics[J]. *Physical Review*, 1957, 106(4): 620–630.
- [19] Shannon C E. A mathematical theory of communication[J]. *The Bell System Technical Journal*, 1948, 27(3): 379–423.
- [20] 张劳模, 庞丽峰, 许等平, 等. 基于最大熵模型预测东北地区红松潜在分布[J]. *江西农业大学学报*, 2020, 42(1): 74–83.
Zhang Laomo, Pang Lifeng, Xu Dengping, et al. Potential distribution of *Pinus koraiensis* in northeastern China predicted by the MaxEnt model[J]. *Acta Agriculturae Universitatis Jiangxiensis*, 2020, 42(1): 74–83.
- [21] 杨庭潇, 马力. 基于最大熵原理的四川省干旱灾害致灾危险性研究[J]. *高原山地气象研究*, 2021, 41(3): 103–107.
Yang Tingxiao, Ma Li. Research on the disaster hazard of drought disasters in Sichuan Province based on the principle of maximum entropy[J]. *Plateau and Mountain Meteorology Research*, 2021, 41(3): 103–107.
- [22] 熊东阳, 张林, 李国庆. 基于最大熵模型的遥感土地利用多分类研究[J]. *自然资源遥感*, 2023, 35(2): 140–148.
Xiong Dongyang, Zhang Lin, Li Guoqing. MaxEnt-based multi-class classification of land use in remote sensing image interpretation[J]. *Remote Sensing for Natural Resources*, 2023, 35(2): 140–148.
- [23] Phillips S J, Dudík M, Schapire R E. Maxent software for modeling species niches and distributions (Version 3.4. 1)[EB/OL]. [2023–07–05]. http://biodiversityinformatics.amnh.org/open_source/maxent/.
- [24] 冯志萍, 余为, 陈新军, 等. 基于最大熵模型的智利外海竹筴鱼栖息地研究[J]. *中国水产科学*, 2021, 28(4): 431–441.
Feng Zhiping, Yu Wei, Chen Xinjun, et al. Distribution of Chilean jack mackerel (*Trachurus murphyi*) habitats off Chile based on a maximum entropy model[J]. *Journal of Fishery Sciences of China*, 2021, 28(4): 431–441.
- [25] 殷晓洁, 周广胜, 隋兴华, 等. 蒙古栎地理分布的主导气候因子及其阈值[J]. *生态学报*, 2013, 33(1): 103–109.
Yin Xiaojie, Zhou Guangsheng, Sui Xinghua, et al. Dominant climatic factors of *Quercus mongolica* geographical distribution and their thresholds[J]. *Acta Ecologica Sinica*, 2013, 33(1): 103–109.
- [26] Phillips S J, Dudík M. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation[J]. *Ecography*, 2008, 31(2): 161–175.

- [27] 刘丹, 李玉堂, 洪玲霞, 等. 基于最大熵模型的吉林省主要天然林潜在分布适宜性[J]. 林业科学, 2018, 54(7): 1–15.
Liu Dan, Li Yutang, Hong Lingxia, et al. The suitability of potential geographic distribution of natural forest types in Jilin Province based on maximum entropy models[J]. *Scientia Silvae Sinicae*, 2018, 54(7): 1–15.
- [28] Mohri M. Seasonal changes in bigeye tuna fishing areas in relation to the oceanographic parameters in the Indian Ocean[J]. *Journal of National Fisheries University*, 1999, 47(2): 43–54.
- [29] 范秀梅, 唐峰华, 崔雪森, 等. 基于栖息地指数的西北太平洋日本鲭渔情预报模型构建[J]. 海洋学报, 2020, 42(12): 34–43.
Fan Xiumei, Tang Fenghua, Cui Xuesen, et al. Habitat suitability index for chub mackerel (*Scomber japonicus*) in the Northwest Pacific Ocean[J]. *Haiyang Xuebao*, 2020, 42(12): 34–43.
- [30] 施展. 基于最大熵模型的森林土壤呼吸分布模拟研究[D]. 杭州: 浙江农林大学, 2021.
Shi Zhan. Simulation research on forest soil respiration distribution based on maximum entropy model[D]. Hangzhou: Zhejiang A&F University, 2021.
- [31] 王运生, 谢丙炎, 万方浩, 等. ROC曲线分析在评价入侵物种分布模型中的应用[J]. 生物多样性, 2007, 15(4): 365–372.
Wang Yunsheng, Xie Bingyan, Wan Fanghao, et al. Application of ROC curve analysis in evaluating the performance of alien species' potential distribution models[J]. *Biodiversity Science*, 2007, 15(4): 365–372.
- [32] 蒋帅. 基于AUC的分类器性能评估问题研究[D]. 长春: 吉林大学, 2016.
Jiang Shuai. Researches on performance evaluation of classifier based on AUC[D]. Changchun: Jilin University, 2016.
- [33] Metz C E. Basic principles of ROC analysis[J]. *Seminars in Nuclear Medicine*, 1978, 8(4): 283–298.
- [34] 王茹琳, 李庆, 封传红, 等. 基于MaxEnt的西藏飞蝗在中国的适生区预测[J]. 生态学报, 2017, 37(24): 8556–8566.
Wang Rulin, Li Qing, Feng Chuanhong, et al. Predicting potential ecological distribution of *Locusta migratoria tibetensis* in China using MaxEnt ecological niche modeling[J]. *Acta Ecologica Sinica*, 2017, 37(24): 8556–8566.
- [35] 崔雪森, 唐峰华, 周为峰, 等. 基于支持向量机的西北太平洋柔鱼渔场预报模型构建[J]. 南方水产科学, 2016, 12(5): 1–7.
Cui Xuesen, Tang Fenghua, Zhou Weifeng, et al. Fishing ground forecasting model of *Ommastrephes bartramii* based on support vector machine (SVM) in the Northwest Pacific[J]. *South China Fisheries Science*, 2016, 12(5): 1–7.
- [36] Ismail A I, Morrison E C, Burt B A, et al. Natural history of periodontal disease in adults: findings from the Tecumseh Periodontal Disease Study, 1959–87[J]. *Journal of Dental Research*, 1990, 69(2): 430–435.
- [37] Development Core R Team. R: A Language and Environment for Statistical Computing[M]. Vienna, Austria: R Foundation for Statistical Computing, 2021.
- [38] Freeman E A, Moisen G. PresenceAbsence: an R package for presence absence analysis[J]. *Journal of Statistical Software*, 2008, 23(11): 1–31.
- [39] 刘庆, 杜雨彤, 邢丹, 等. 基于MaxEnt模型对我国三带喙库蚊潜在分布区的预估研究[J]. 寄生虫与医学昆虫学报, 2022, 29(4): 203–211.
Liu Qing, Du Yutong, Xing Dan, et al. Prediction of the potential distribution areas of *Culex tritaeniorhynchus* in China based on Max-Ent model[J]. *Acta Parasitologica et Medica Entomologica Sinica*, 2022, 29(4): 203–211.
- [40] 孔维尧, 李欣海, 邹红菲. 最大熵模型在物种分布预测中的优化[J]. 应用生态学报, 2019, 30(6): 2116–2128.
Kong Weiyao, Li Xinhai, Zou Hongfei. Optimizing MaxEnt model in the prediction of species distribution[J]. *Chinese Journal of Applied Ecology*, 2019, 30(6): 2116–2128.
- [41] 邢丁亮, 郝占庆. 最大熵原理及其在生态学中的应用[J]. 生物多样性, 2011, 19(3): 295–302.
Xing Dingliang, Hao Zhanqing. The principle of maximum entropy and its applications in ecology[J]. *Biodiversity Science*, 2011, 19(3): 295–302.
- [42] Berger A L, Della Pietra V J, Della Pietra S A. A maximum entropy approach to natural language processing[J]. *Computational Linguistics*, 1996, 22(1): 39–71.
- [43] Bovee K D. Development and evaluation of habitat suitability criteria for use in the instream flow incremental methodology[R]. Washington: USDI Fish and Wildlife Service, 1986.
- [44] Franklin J. Mapping Species Distributions: Spatial Inference and Prediction[M]. Cambridge: Cambridge University Press, 2010.
- [45] Elith J, Graham C H, Anderson R P, et al. Novel methods improve prediction of species' distributions from occurrence data[J]. *Ecography*, 2006, 29(2): 129–151.
- [46] Li Wenkai, Guo Qinghua, Elkan C. Can we model the probability of presence of species without absence data?[J]. *Ecography*, 2011, 34(6): 1096–1105.
- [47] Tian Siquan, Chen Xinjun, Chen Yong, et al. Evaluating habitat suitability indices derived from CPUE and fishing effort data for *Ommastrephes bartramii* in the northwestern Pacific Ocean[J]. *Fisheries Research*, 2009, 95(2/3): 181–188.
- [48] 崔雪森, 周灿, 唐峰华, 等. 西北太平洋柔鱼渔场非参数栖息地适宜性指数模型[J]. 广东海洋大学学报, 2020, 40(6): 53–62.
Cui Xuesen, Zhou Can, Tang Fenghua, et al. Nonparametric habitat suitability index model for *Ommastrephes bartramii* fishing ground in the Northwest Pacific Ocean[J]. *Journal of Guangdong Ocean University*, 2020, 40(6): 53–62.
- [49] 晏然, 范江涛, 徐姗姗, 等. 南海北部近海竹荚鱼栖息地分布特征[J]. 生态学杂志, 2018, 37(8): 2430–2435.
Yan Ran, Fan Jiangtao, Xu Shannan, et al. Distribution characteristics of jack mackerel (*Trachurus japonicus*) habitat in the offshore wa-

ters of northern South China Sea[J]. Chinese Journal of Ecology, 2018, 37(8): 2430–2435.

- [50] 易明华, 官文江, 陈新军. 基于理想自由分布理论对CPUE与渔业资源关系的探讨——以我国近海鲈灯光围网渔业为例[J]. 大连水产学院学报, 2009, 24(4): 325–330.

Yi Minghua, Guan Wenjiang, Chen Xinjun. The relationship between CPUE and fish abundance based on ideal free distribution theory: take the large light purse seine fishery of mackerel in Yellow Sea and East China Sea as an example[J]. Journal of Dalian Fisheries University, 2009, 24(4): 325–330.

Prediction of chub mackerel fishing ground distribution in the East China Sea and Yellow Sea based on maximum entropy model and habitat suitability index model

Cao Ruixing¹, Guan Wenjiang^{1,2}, Gao Feng^{1,3}, He Weiwei¹

(1. College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China; 2. The Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources, Ministry of Education, Shanghai 201306, China; 3. National Distant-water Fisheries Engineering Research Center, Shanghai Ocean University, Shanghai 201306, China)

Abstract: The maximum entropy model (Maxent) and habitat suitability index (HSI) model are widely used in fishery forecasting studies. To compare the forecasting performance of these two models on fishing grounds and improve the scientific management of chub mackerel (*Scomber japonicus*) resources, this study used the fishery data of chub mackerel in the East China Sea and Yellow Sea from 2003 to 2012, and marine environmental data, including sea surface temperature, sea surface height, sea surface salinity and sea surface temperature gradient, to construct the Maxent model and HSI model. The aim was to analyze and compare the effectiveness of these two models in predicting the habitat of chub mackerel in the East China Sea and Yellow Sea. The quantitative evaluation of the prediction performance of the two models was conducted using the area under curve (AUC) of the receiver operating characteristic (ROC), and the correspondence between the probability of fishing grounds predicted by the models and the percentage of the actual catches. The results showed that: (1) locations predicted by the maximum entropy model to have a high probability of fishing occurrence coincided with actual fishing locations. The probability of predicting fishery occurrence in the sea area without historical fishing data was lower. Locations predicted to have a high habitat index by the HSI model partially overlapped with actual fishing locations. A high habitat index was obtained in the sea area without historical fishing data. The probability of the HSI model predicting non-fishing grounds as fishing grounds was higher than that of the Maxent model; (2) the monthly average AUC values of the Maxent and HSI model were 0.95 and 0.66, respectively, indicating that the Maxent had relatively better predictive results; (3) when using the HSI model, non-fishing grounds data should be added to the model, and the collection of such data should be strengthened otherwise, there is a possibility of overestimation when such models forecast fishing grounds. When using the Maxent, the spatial coverage of fishery data must be improved otherwise, it cannot fully reflect the spatial and temporal distribution dynamics of the fishery. The results of this study provide a reference for improving the accuracy of forecasting for the chub mackerel fishery in the East China Sea and Yellow Sea.

Key words: maximum entropy model; habitat suitability index model; model comparison; chub mackerel