

张涛, 赵天亚, 栾静, 等. 基于物种分布模型分析环境因子对海州湾偶见种资源分布的影响[J]. 海洋学报, 2023, 45(7): 69–78, doi:10.12284/hyxb2023078

Zhang Tao, Zhao Tianya, Luan Jing, et al. Analysis of the influence of environmental factors on the distribution of occasional species in the Haizhou Bay based on species distribution model[J]. Haiyang Xuebao, 2023, 45(7): 69–78, doi:10.12284/hyxb2023078

# 基于物种分布模型分析环境因子对海州湾 偶见种资源分布的影响

张涛<sup>1,2</sup>, 赵天亚<sup>1,2</sup>, 栾静<sup>1,2</sup>, 张云雷<sup>1,2</sup>, 张崇良<sup>1,2\*</sup>

(1. 中国海洋大学 水产学院, 山东 青岛 266003; 2. 海州湾渔业生态系统教育部野外科学观测研究站, 山东 青岛 266003)

**摘要:** 偶见种易受环境变化和人类活动等外界威胁, 在生物多样性保护中具有重要参考价值, 但由于其数据量较少、分析困难, 目前对分布特征的研究较少, 其分布与环境因子的关系尚待探究。本研究基于 2013–2019 年海州湾渔业资源调查数据, 分析了凤鲚 (*Coilia mystus*)、红狼牙虾虎鱼 (*Odontamblyopus rubicundus*) 和蛇鲻 (*Erisphex pottii*) 3 种海州湾偶见种资源分布与环境因子的关系, 并比较了广义可加模型 (GAM) 和随机森林 (RF) 模型对其资源分布的拟合效果, 采用交叉验证的方法对模型的预测性能进行了评价。结果显示, 水深是影响春、秋季凤鲚和红狼牙虾虎鱼资源分布的最显著因子, 而底层水温仅在秋季是影响蛇鲻资源分布的最重要环境因子。凤鲚分布模型的方差解释率最高, 其次为红狼牙虾虎鱼, 蛇鲻模型方差解释率最低。凤鲚、红狼牙虾虎鱼和蛇鲻分布模型在春季方差解释率均低于秋季。交叉验证表明, 3 个物种预测结果的曲线下面积 (AUC) 值在 0.70~0.85 之间, 仅秋季凤鲚的 AUC 值达到 0.9; 同时 GAM 预测结果的 AUC 值均大于 RF 模型, 表明对于偶见种而言, GAM 的预测性能优于 RF 模型。本研究为今后开展偶见种研究的模型选择提供了参考, 对偶见种资源保护具有指导意义。

**关键词:** 海州湾; 偶见种; 广义可加模型 (GAM); 随机森林 (RF) 模型; 交叉验证

中图分类号: S932.4; P735

文献标志码: A

文章编号: 0253-4193(2023)07-0069-10

## 1 引言

偶见种是指群落中个体数或出现频率较少的种类, 容易受到外界威胁而处于濒危或灭绝状态<sup>[1]</sup>。群落中往往有较多的种类均属于偶见种<sup>[2]</sup>, 但每个物种的数据量较少、分析困难。在群落研究中往往只关注数量较多、优势度较大的优势种<sup>[3-4]</sup>和常见种等, 而对于偶见种的研究比较少见。然而, 偶见种是群落生物多样性的重要组分, 对于群落的稳定性具有重要作用, 对群落和生态系统中不同过程产生影响。因此,

对偶见种栖息习性和资源状况的研究具有重要意义, 可以为保护偶见种资源和生物多样性提供参考, 但目前对于偶见种的研究多集中于植物学领域<sup>[5-7]</sup>, 海洋学领域对于偶见种的研究相对较少。

了解偶见种的栖息分布规律是对其开展保护和管理的基礎。当前, 物种分布模型 (Species Distribution Model, SDM)<sup>[8]</sup> 已成为生态学和应用保护生物学的重要方法<sup>[9]</sup>。随着技术的发展, 基于回归模型和机器学习方法的物种分布模型在渔业领域的应用越来越广泛, 应用物种分布模型分析物种的资源分布与环

收稿日期: 2022-09-23; 修订日期: 2023-02-20。

基金项目: 国家重点研发计划 (2018YFD0900904, 2018YFD0900906)。

作者简介: 张涛 (1998—), 男, 河南省商丘市人, 主要从事渔业资源研究。E-mail: zhangtao9807@163.com

\* 通信作者: 张崇良, 副教授, 主要从事渔业资源评估、生态系统模拟。E-mail: zhangclg@ouc.edu.cn

境因子的关系<sup>[10]</sup>更是研究的热点领域,如崔晏华等<sup>[11]</sup>利用随机森林(Random Forest, RF)模型分析了海州湾春季短蛸(*Octopus ocellatus*)的栖息分布与环境因子的关系,纪毓鹏等<sup>[12]</sup>使用广义可加模型(Generalized Additive Model, GAM)分析了山东南部近海脊腹褐虾(*Crangon affinis*)的时空分布与水深、海水底层温度等环境因子间的关系。然而,受数据量的限制,针对偶见种的物种分布建模存在一定的困难,常见模型对偶见种的预测效果尚缺乏检验。有研究关注了偶见种建模时的过度拟合问题,如 Lomba 等<sup>[13]</sup>为解决稀少物种预测建模过程存在的“稀有物种建模悖论”,提出了一种针对稀有种的建模方法,以避免统计模型的过度拟合; Breiner 等<sup>[14]</sup>为克服稀少物种建模的局限性,开发了一种使用小型模型集合的新策略,其表现出较传统物种分布模型更优越的性能。

为了深入探究 SDM 建模方法对于偶见种栖息分布分析的影响,本研究根据 2013–2019 年在海州湾进行的 14 个航次的渔业资源底拖网调查数据,采用 GAM 和基于机器学习的 RF 模型,分析了海州湾海域 3 种偶见种——凤鲚(*Coilia mystus*)、红狼牙虾虎鱼(*Odontamblyopus rubicundus*)和虹鲃(*Erisphex pottii*)的资源分布与环境因子的关系。研究采用交叉验证的方法对模型进行了评价,检验了渔业领域两种广泛应用的模型对于偶见种的拟合效果和预测性能。本研究结果可为偶见种的资源保护提供参考,为物种分布模型在偶见种方面的应用提供科学依据。

## 2 材料与方法

### 2.1 数据来源

在海州湾海域采用分层随机取样的方法设计调查站位,调查范围为 34°20′~35°40′N, 119°20′~121°10′E。根据水深等环境特征将调查海域分为 A–E 共 5 个采样区(图 1),并在每区按经纬度划分为 10′×10′的采样小区。2013–2019 年春季(4、5 月)、秋季(9、10 月)进行了 14 个航次的渔业资源底拖网调查,每个航次在各区域内随机选取约 18 个调查站位,7 年间共调查 248 个站位。

调查船为单拖渔船,功率为 220 kW,拖速在 2~3 km 之间,每站拖网时间约为 1 h。拖网时网口水平扩张宽度约为 25 m,网囊网目约为 17 mm。在每个调查站位捕获渔获物的同时,同步使用 CTD 温盐深仪进行相关环境参数的测定,包括水深、水温、盐度和叶绿素 *a* 浓度。样品的采集与处理均按照《海洋调查规范》(GB/T 1273.6–2007)<sup>[15]</sup>进行。

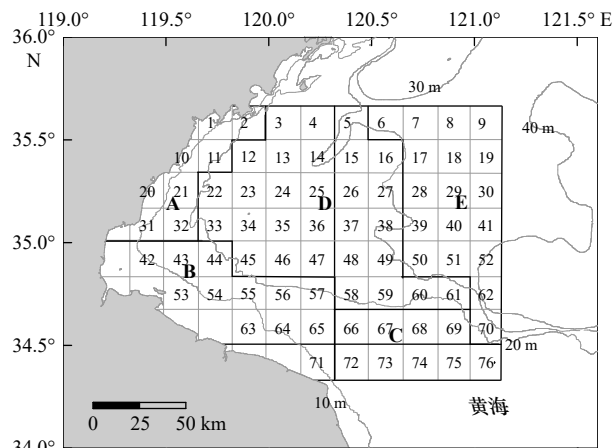


图 1 海州湾调查区域

Fig. 1 Sampling areas in the Haizhou Bay

灰色为等深线

The gray line is the isobath line

### 2.2 模型构建

本研究聚焦海州湾偶见种,根据前期对海州湾偶见种的相关调查结果<sup>[16]</sup>,选取凤鲚、红狼牙虾虎鱼和虹鲃作为本研究的目标物种,其相对重要性指数(Index of Relative Importance, IRI)分别为 2.73、1.67、4.06,较为少见。将偶见种的出现概率作为建模响应变量。若某站位捕获偶见种,其观测值取为 1,未捕获则为 0,利用偶见种出现/不出现的数据建立物种分布模型。在渔业领域中,物种分布模型在研究物种栖息分布方面有着较为广泛的应用,根据相关研究<sup>[17–19]</sup>,本文采用两种应用广泛的建模方法——GAM 和 RF 模型开展分析,评估其对于偶见种的预测性能。

#### 2.2.1 广义可加模型

GAM 是广义线性模型的非线性拓展,是渔业领域最常用的物种分布模型之一,其特点在于可以通过引入平滑函数来反映响应变量与解释变量之间的非线性关系<sup>[20]</sup>,模型由随机部分、系统(叠加)部分和联系两者的链接函数 3 部分组成<sup>[21]</sup>。本文中的 GAM 假设误差服从二项分布,取值为 0 或 1 时,方程左右两侧不能对应,为继续线性回归,采用 logistic 链接函数进行变换,分析偶见种的出现概率与环境因子的关系,其一般表达式为

$$\text{logit}(P) = \alpha + f_1(X_1) + f_2(X_2) + \dots + \varepsilon, \quad (1)$$

式中,  $\text{logit}(\ )$  为链接函数;  $P$  为偶见种在此站点出现的概率;  $\alpha$  为适合函数的截距;  $X$  为解释变量,即影响因子;  $f(X)$  为样条平滑函数;  $\varepsilon$  为随机误差项。

#### 2.2.2 随机森林模型

RF 模型是一种基于分类回归树算法的机器学习算法<sup>[22]</sup>,其原理是利用 bootstrap 重抽样方法从原始样

本中随机抽取多个样本,通过这种方式生成成百上千棵分类回归树,然后综合分类回归树的预测,以投票或取均值的方法得出最终预测结果<sup>[23-24]</sup>。RF模型不仅对具有多维度变量的大数据处理较快,且具有较强的适应性,同时具有很高的分类精度、预测精度和较强的泛化能力<sup>[25]</sup>,不容易出现过拟合的现象,对量化复杂的非线性关系有较好的效果。

### 2.3 因子筛选与模型拟合

本文选取水深、表层温度(SST)、底层温度(SBT)、表层盐度(SSS)和底层盐度(SBS)作为主要环境因子,研究表明<sup>[16]</sup>,海州湾偶见种在经度方向上的空间分布特征更为明显,因此选择经度作为空间因子进行建模分析。应用方差膨胀系数(Variance Inflation Factor, VIF)<sup>[26]</sup>对春、秋季的影响因子分别进行多重共线性检验,筛选适合加入模型的因子, VIF > 4的因子在建模之前予以去除。

在模型构建中,利用逐步回归法<sup>[27]</sup>进行因子的筛选,依照赤池信息准则(Akaike Information Criterion, AIC)<sup>[28]</sup>进行最优模型的建立,即在AIC值最小的模型中不断加入新的因子,直至所有因子添加完毕,从中筛选AIC最小的模型即为最优模型。AIC的一般表达式为

$$AIC = 2k - 2 \ln L, \quad (2)$$

式中,  $k$  为参数的个数;  $L$  为似然函数。

应用AIC和方差解释率来检验和比较模型的拟合效果,即AIC值越小,累计方差解释率越大,则模型的拟合效果越好。

### 2.4 模型预测性能评估

本文使用交叉验证法来检验和比较GAM和RF模型的预测性能。交叉验证过程首先在全部数据中随机选择60%的数据,作为训练数据建立模型,剩余40%的数据作为检验数据来评估模型的性能。使用受试者操作特征(Receiver Operating Characteristic, ROC)<sup>[29]</sup>曲线下面积(Area Under The Curve, AUC)作为模型预测性能的评价指标,其取值范围为0~1<sup>[30]</sup>。AUC值越大说明模型的预测性能越好,反之则越差,0表示模型的性能最差,无法进行有效预测;1表示模型的性能最好,存在至少一个阈值能得出完美预测,但一般来说不可能实现。将上述交叉验证过程重复100次,取平均值即为最终的AUC值,表征模型对偶见种预测的准确性。

本文中的模型构建和评估过程均通过R 4.0.3软件来完成,利用“mgcv”包进行GAM的构建,利用“randomForest”包进行RF模型的构建,利用“pROC”

包进行模型预测性能评估,利用“ggplot2”包绘制交叉验证结果图。

## 3 结果

### 3.1 因子筛选

应用VIF对春、秋季的影响因子分别进行多重共线性检验,将VIF大于4的表层温度、表层盐度去除,剩余因子的VIF均小于4(表1),检验后得到可以加入模型的因子包括:水深、底层温度、底层盐度和经度。

表1 海州湾春、秋季各环境因子的方差膨胀系数  
Table 1 Variance inflation factor of each environmental factor during spring and autumn in the Haizhou Bay

季节	水深	底层温度	底层盐度	经度
春季	2.29	1.58	1.67	1.41
秋季	1.83	1.17	1.62	1.52

### 3.2 模型构建

根据模型AIC和方差解释率的标准,GAM和RF模型分别筛选出最适因子,组成最优模型(表2)。在春季,凤鲚最优GAM的解释变量为水深、经度和底层盐度,其AIC值最小为69.61,方差解释率为38.9%。RF模型的方差解释率偏低,为18.0%。秋季,凤鲚最优GAM的解释变量为水深、经度和底层温度,其AIC值为51.30,方差解释率为51.2%。RF模型的方差解释率偏低,为30.1%。秋季凤鲚GAM和RF模型的方差解释率均较春季更大。

在春季,红狼牙虾虎鱼最优GAM的解释变量为水深、底层盐度和底层温度,其AIC值最小为83.03,方差解释率为25.1%。RF模型的方差解释率偏低,为5.0%。秋季,红狼牙虾虎鱼最优GAM的解释变量为水深、经度和底层盐度,其AIC值最小为47.79,方差解释率为46.2%。RF模型的方差解释率偏低,为7.7%。同样,秋季红狼牙虾虎鱼GAM和RF模型的方差解释率较春季更大。

在春季,虹鲃最优GAM的解释变量为水深和底层温度,其AIC值最小为69.63,方差解释率为18.2%。RF模型的方差解释率偏低,为7.5%。秋季,虹鲃最优GAM的解释变量为底层温度、底层盐度和经度,其AIC值最小为100.31,方差解释率为34.1%。RF模型的方差解释率偏低,为26.4%。秋季虹鲃GAM和RF模型的方差解释率也均大于春季。

### 3.3 预测性能

交叉验证表明,春季凤鲚、红狼牙虾虎鱼和虹鲃

表 2 春、秋季 3 个偶见种最优模型

Table 2 Optimal model for three occasional species during spring and autumn

季节	物种	模型	解释变量	AIC	方差解释率/%
春季	凤鲚 <i>C. mystus</i>	GAM	Depth+Longitude+SBS	69.61	38.9
		RF	Depth+SBT+Longitude		18.0
	红狼牙虾虎鱼 <i>O. rubicundus</i>	GAM	Depth+SBS+SBT	83.03	25.1
		RF	Depth+SBT+SBS		5.0
	虹鲷 <i>E. pottii</i>	GAM	Depth+SBT	69.63	18.2
		RF	Depth+SBS+SBT+Longitude		7.5
秋季	凤鲚 <i>C. mystus</i>	GAM	Depth+Longitude+SBT	51.30	51.2
		RF	Depth+Longitude		30.1
	红狼牙虾虎鱼 <i>O. rubicundus</i>	GAM	Depth+Longitude+SBS	47.79	46.2
		RF	Longitude+SBS+Depth		7.7
	虹鲷 <i>E. pottii</i>	GAM	SBT+SBS+Longitude	100.31	34.1
		RF	SBT+SBS+Longitude+Depth		26.4

注: Depth、SBT、SBS和Longitude分别代表环境因子水深、底层温度、底层盐度和经度。

GAM 的 AUC 值均大于 RF 模型(图 2), 分别为  $0.85 \pm 0.05$ 、 $0.78 \pm 0.08$  和  $0.77 \pm 0.10$ 。秋季凤鲚和虹鲷 GAM

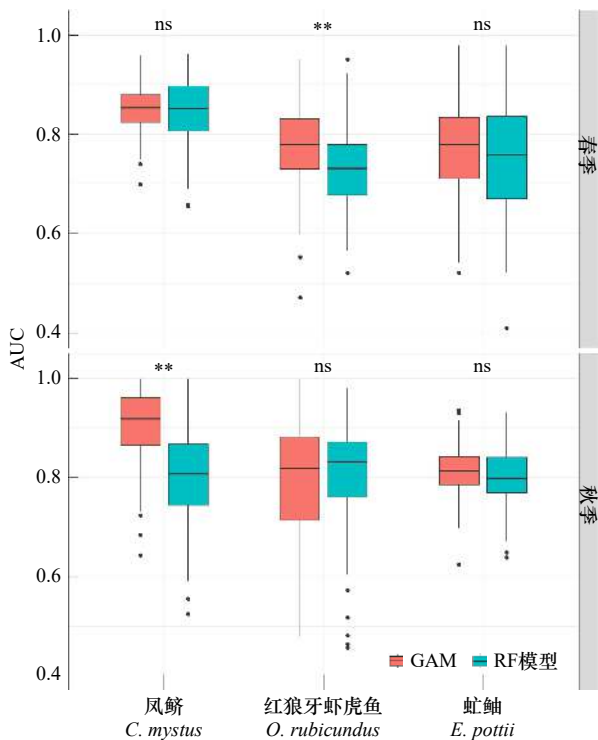


图 2 3 种偶见种 GAM 与 RF 模型交叉验证结果

Fig. 2 Cross-validation of GAM and RF model for three occasional species

ns 代表差异不显著 ( $p > 0.05$ ); \*\*代表差异极显著 ( $p < 0.01$ )

ns represent no significant difference ( $p > 0.05$ ); \*\* represent extremely significant difference ( $p < 0.01$ )

的 AUC 值大于 RF 模型, 分别为  $0.90 \pm 0.08$  和  $0.81 \pm 0.05$ , 而红狼牙虾虎鱼 GAM 的 AUC 值小于 RF 模型, 其 GAM 和 RF 模型的 AUC 值分别为  $0.79 \pm 0.12$  和  $0.81 \pm 0.11$ 。t 检验结果表明, 秋季凤鲚和春季红狼牙虾虎鱼 GAM 的 AUC 值均显著大于 RF 模型 ( $p < 0.01$ ), 而其余模型比较均不显著。

### 3.4 偶见种分布与环境因子的关系

由于 GAM 的拟合和预测性能较 RF 模型更好, 因此以 GAM 最优模型分析偶见种资源分布与环境因子的关系。

在春、秋两季, 凤鲚的分布与水深因子显著相关(图 3)。春季时, 在 20 m 以浅, 其出现概率随水深的变化较为平缓, 水深大于 20 m, 出现概率随水深的增加呈现较为明显的下降; 秋季其出现概率随水深的增加呈现直线下降的趋势。凤鲚分布其次受到经度的影响, 出现概率随经度的增大而缓慢上升。在春季, 凤鲚的出现概率随盐度的变化呈现先升高后下降的趋势; 秋季, 凤鲚的出现概率与底层温度呈现较高相关性, 随底层温度的增加, 其出现概率也逐渐上升。

春、秋季红狼牙虾虎鱼的出现概率均随水深的增加呈现减小的趋势。春季其出现概率随底层盐度和底层温度的增加呈现一直减小的状态, 而在秋季经度成为影响红狼牙虾虎鱼的出现概率的重要因子, 其出现概率在  $120.6^{\circ}\text{E}$  达到最大; 同时, 随底层盐度先上升, 在 29.5 后开始下降(图 4)。

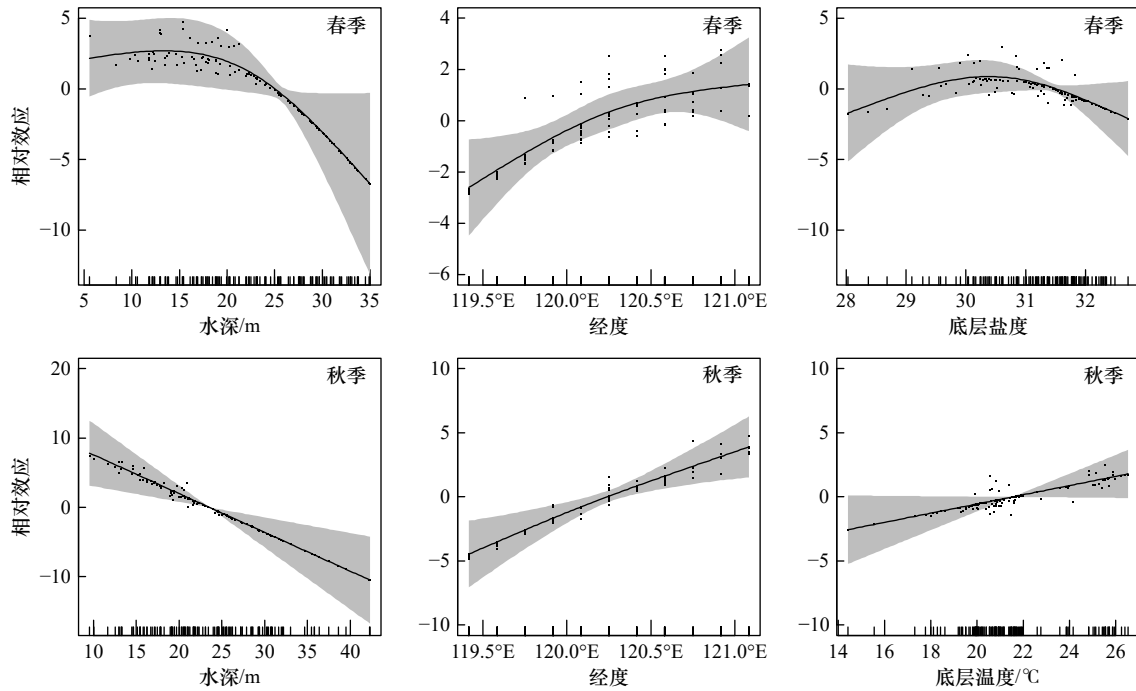


图3 GAM中环境因子对凤鲚出现概率的影响

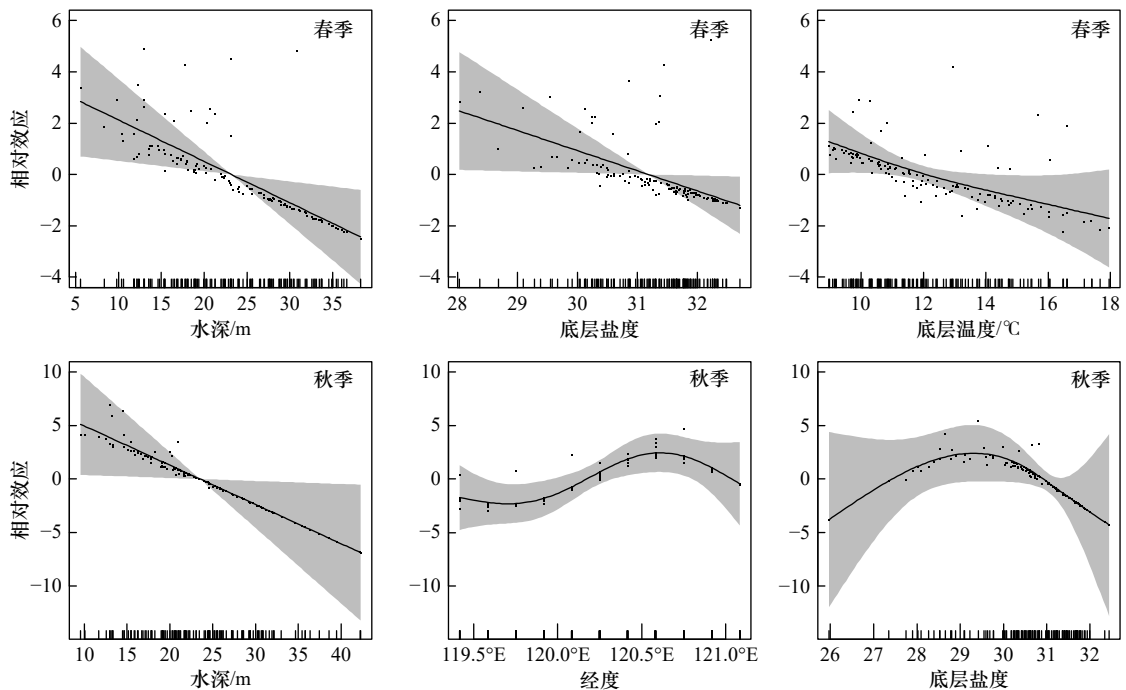
Fig. 3 Influence of environmental factors on the occurrence probability of *Coilia mystus* in the GAM

图4 GAM中环境因子对红狼牙虾虎鱼出现概率的影响

Fig. 4 Influence of environmental factors on the occurrence probability of *Odontamblyopus rubicundus* in the GAM

影响虹鲃春、秋季分布的环境因子差别较大(图5)。在春季,虹鲃的分布与水深因子显著相关,其出现概率随水深的增加呈现上升的趋势;其次为底层温度,其出现概率随底层温度的增加而上升。在秋季,虹鲃的分布受底层温度的影响最显著,在19℃后随底层温度的增加迅速下降。秋季虹鲃的分布还与底层盐度和经度具有较高相关性,在底层盐度达到31.0前,

其出现概率呈现上升的趋势,然后开始小幅下降;其出现概率随经度的增加而上升,变化较为平缓。

## 4 讨论

### 4.1 模型比较

本文针对凤鲚、红狼牙虾虎鱼和虹鲃3种偶见种,分别比较了其GAM和RF模型的拟合效果和预测性

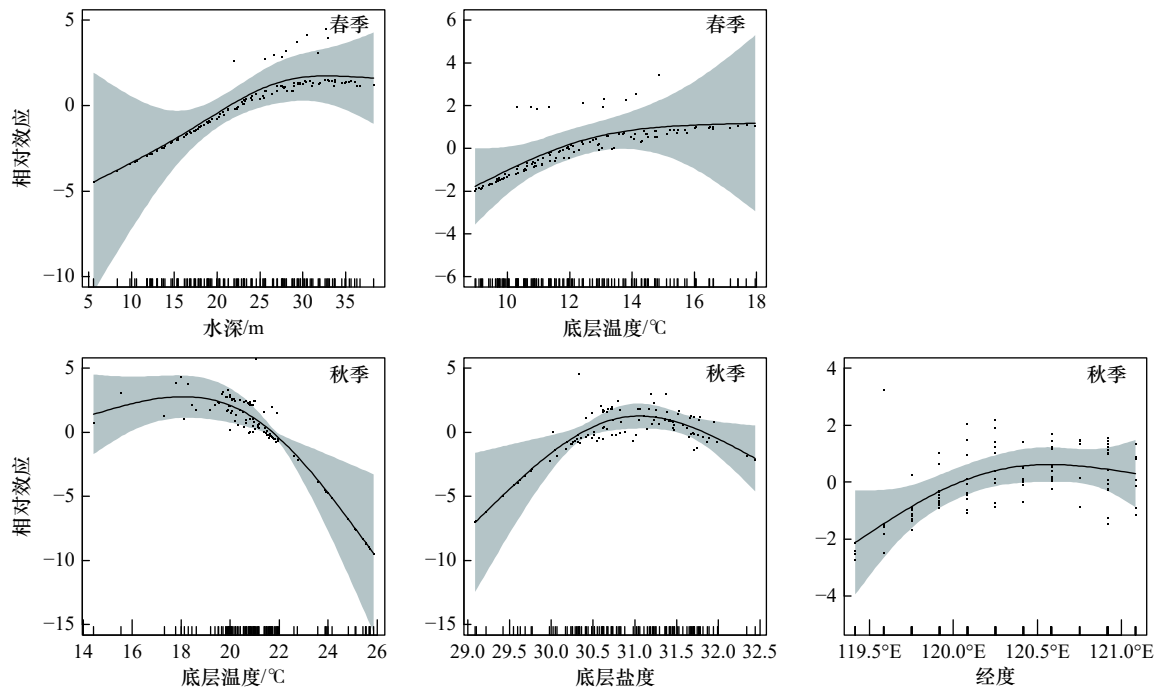


图5 GAM中环境因子对虹鲆出现概率的影响

Fig. 5 Influence of environmental factors on the occurrence probability of *Erisphex pottii* in the GAM

能。从整体来看, 凤鲚、红狼牙虾虎鱼和虹鲆的 GAM 最优模型的方差解释率均大于 RF 最优模型的方差解释率, 红狼牙虾虎鱼的方差解释率差距尤为明显, 在春、秋两季 GAM 最优模型的方差解释率分别为 25.1% 和 46.2%, RF 最优模型的方差解释率则分别为 5.0% 和 7.7%, GAM 的方差解释率均远远大于 RF 模型。因此, 可以认为在偶见种模型构建中, GAM 的拟合效果较 RF 模型更好, 这与栾静等<sup>[31]</sup> 和倪一卓等<sup>[32]</sup> 的研究结果一致。其原因可能是, 作为一种传统的物种分布模型, GAM 通过建立回归方程来表示分布与环境因子的关系, 其定义的平滑函数可能使模型更好地服从偶见种的存在-不存在数据结构, 模型本身的解释能力较好<sup>[20]</sup>; 另一方面, GAM 应用 AIC 准则构建最优模型, 这可能会高估参数数量的影响<sup>[31]</sup>, 导致 GAM 可能会选择更复杂的模型<sup>[33]</sup>, 需要注意 GAM 可能会对数据出现过拟合的现象。RF 模型的原理是对数据和变量进行随机筛选, 生成很多的分类回归树, 最后汇总分类回归树的结果来得到结论, 其优点是可以在很大程度上提高模型的预测性能, 同时避免模型过度拟合, 但缺点是可能会降低模型对训练数据集的拟合效果, 降低其模型解释率。综合以上两点, 我们认为在数据量较少的偶见种的物种分布模型研究中, 广义可加模型的拟合效果会强于随机森林模型。

本文中交叉验证得到 AUC 值来表征模型的预测性能, 其 AUC 值多在 0.70~0.85 之间, 仅秋季凤

鲚 GAM 的 AUC 值达到 0.9。AUC 的数值越大, 表征预测结果越精确, 其中 0.7~0.8 表明模型预测性能一般, 0.8~0.9 表明模型预测性能良好, 0.9~1.0 表明模型预测性能优秀<sup>[34]</sup>。因此对于偶见种而言, 传统 SDM 的预测性能尚有待提升。此外, 除秋季红狼牙虾虎鱼 GAM 的 AUC 值略小于 RF 模型, 其余物种的 GAM 的 AUC 值均略高于 RF 模型(图 2), 且  $t$  检验结果显示, 秋季凤鲚和春季红狼牙虾虎鱼 GAM 的 AUC 值均显著大于 RF 模型( $p < 0.01$ )。这表明在数据量较少的偶见种的物种分布模型研究中, GAM 的预测性能强于 RF 模型, 这与栾静等<sup>[31]</sup> 和 Li 等<sup>[35]</sup> 针对常见种的研究结论不符。物种分布模型受数据观测误差的影响较大, 由于偶见种的特点, 在数据量较少的情况下测量误差可能会较大, 从而影响建模效果。有研究表明, 如果不能很好地控制测量误差, 机器学习的预测性能甚至不如常见的传统统计模型<sup>[36]</sup>, 这与本研究的结论一致。因此, 本研究认为由于偶见种本身的数据量较小, 观测误差可能是导致广义可加模型的预测性能强于随机森林模型的主要原因。

#### 4.2 环境因子的影响

本研究显示, 凤鲚的分布主要受到水深的影响, 主要在水深小于 40 m 的海域中出现, 且春季主要分布在水深小于 35 m 的海域中, 在水深 20 m 的海域出现概率达到峰值(图 3), 这可能与短距离洄游产卵的习性有关。凤鲚是一种河口性的洄游鱼类, 以浮游

生物为食, 而海州湾丰富的饵料资源可为其产卵和生长提供丰富的物质基础。凤鲚每年春季都会从海中向河口区域洄游产卵<sup>[37]</sup>, 这可能导致春季凤鲚在35 m以深的海域没有捕获。此外, 盐度对鱼类的分布有重要的影响<sup>[38]</sup>, 在本文中, 春季凤鲚在底层盐度30~32.5之间的出现概率最大, 据此推测春季凤鲚在海州湾洄游产卵期的最适宜底层盐度范围为30~32.5。

水深对春、秋季红狼牙虾虎鱼的分布均有明显的影响(图4), 其出现概率与水深呈负相关关系。红狼牙虾虎鱼喜爱栖息于浅海及河口附近泥沙中, 常在泥沙中钻穴营居<sup>[39]</sup>, 这与本文中红狼牙虾虎鱼分布于10~35 m水深海域的结论一致。此外, 底层盐度也是影响红狼牙虾虎鱼分布的重要条件, 本研究表明海州湾红狼牙虾虎鱼在春、秋季的适盐范围大致相同, 在30~32之间。目前国内对红狼牙虾虎鱼的研究较少, 对海州湾此类鱼种的研究更是匮乏, 该结论还待进一步验证。

影响虹鲃春、秋季分布的环境因子差异较大(图5), 春季主要受到水深的影响, 其出现概率与水深呈正相关关系, 这也符合虹鲃喜爱栖息于较深海域的习性。而春、秋季底层水温对虹鲃的影响则差异较大, 春季虹鲃栖息于较深的海区, 其栖息水温较低, 适宜的底层水温在9~16℃之间; 但在秋季, 底层水温取代水深成为影响虹鲃分布的显著因素, 夏、秋季海州湾的温度较春季高, 因此秋季虹鲃的适宜水温较春季更高, 在秋季十分集中地分布在19~22℃之间, 据此推测秋季温度的变化对于虹鲃种群动态具有重要影响, 在生物多样性的保护当中应予以关注。

比较发现, 各偶见种在秋季最优模型的方差解释率均较春季更大, 这表明秋季环境因子对其资源分布的影响更显著。同时, 春、秋两季中各偶见种不同模型对环境因子的选择均较为相似(表2), 除秋季虹鲃的主要因子为底层水温外, 其余季节凤鲚、红狼牙虾

虎鱼和虹鲃中各最优模型筛选的主要影响因子均为水深。需要说明的是, 水深因子作为一种综合性的环境因子<sup>[40]</sup>, 它可以通过影响温度和盐度等因子而间接影响生物的分布。在本研究中, 水深因子作为模型中的主要因子, 更多是直接反映了凤鲚和红狼牙虾虎鱼适宜生活在浅海的生活习性。

## 5 结论

物种分布模型在生态研究和生物保护方面发挥了重要作用, 但每个模型各有优缺点和适用对象<sup>[41]</sup>。为了探究不同情景下的模型适用情况, 模型间的比较研究也越来越受到关注<sup>[42]</sup>。保护生物学家对稀有物种更为关注, 探索保护和修复的科学方法<sup>[43]</sup>, 但这些物种通常具有有限的可用数据, 为准确选择物种分布模型带来了挑战<sup>[44]</sup>。本研究聚焦于海州湾偶见种, 比较了两种模型对偶见种的预测效果, 表明传统的GAM和RF模型在偶见种的预测方面仍存在一定局限。由于偶见种个体数或出现频率较小, 导致其0值较多, 因此本文采用0~1数据模式即缺失/存在数据模式, 在本研究中, 缺失0和存在1数据均作为有效数据参与建模, 但大量0值的存在可能会导致较大的测量误差, 从而影响模型的结果<sup>[36]</sup>, 因此未来研究偶见种的建模还需探索更多物种分布模型的适用性, 如两阶段广义可加模型(two-stage Generalized Additive Model)<sup>[45]</sup>、Tweedie-GAM<sup>[46]</sup>和最大熵模型(MaxEnt Model)均对0值稍多的物种表现出相对较好的模型效果, 但偶见种数据中的0值占比远大于上述研究物种, 所以此类模型对于偶见种的使用效果尚不得而知, 未来可以进一步展开研究。随着人类活动和气候变化等的影响, 对稀有种、偶见种和生物多样性的研究保护越来越受到重视, 而本文作为一个案例, 可为偶见种的资源保护以及建立海洋保护区提供参考。未来需要探究更多模型以改进偶见种的建模方法, 在更大的空间和时间尺度上开展偶见种栖息分布研究。

## 参考文献:

- [1] 朱靖. 稀有种及其保护[J]. 野生动物保护, 1979(1): 1-8, 28.  
Zhu Jing. Rare species and their protection[J]. Chinese Journal of Wildlife, 1979(1): 1-8, 28.
- [2] Trindade-Santos I, Moyes F, Magurran A E. Global patterns in functional rarity of marine fish[J]. *Nature Communications*, 2022, 13(1): 877.
- [3] 李显森, 于振海, 孙珊, 等. 长江口及其毗邻海域鱼类群落优势种的生态位宽度与重叠[J]. 应用生态学报, 2013, 24(8): 2353-2359.  
Li Xiansen, Yu Zhenhai, Sun Shan, et al. Ecological niche breadth and niche overlap of dominant species of fish assemblage in Changjiang River Estuary and its adjacent waters[J]. *Chinese Journal of Applied Ecology*, 2013, 24(8): 2353-2359.
- [4] 王雪辉, 杜飞雁, 邱永松, 等. 1980-2007年大亚湾鱼类物种多样性、区系特征和数量变化[J]. 应用生态学报, 2010, 21(9): 2403-2410.  
Wang Xuehui, Du Feiyan, Qiu Yongsong, et al. Variations of fish species diversity, faunal assemblage, and abundances in Daya Bay in

- 1980–2007[J]. Chinese Journal of Applied Ecology, 2010, 21(9): 2403–2410.
- [5] 刘旻霞, 李全弟, 蒋晓轩, 等. 甘南亚高寒草甸稀有种对物种多样性和物种多度分布格局的贡献[J]. 生物多样性, 2020, 28(2): 107–116.
- Liu Minxia, Li Quandi, Jiang Xiaoxuan, et al. Contribution of rare species to species diversity and species abundance distribution pattern in the Gannan subalpine meadow[J]. Biodiversity Science, 2020, 28(2): 107–116.
- [6] 周文嵩. 海南铜鼓岭热带常绿季雨矮林偶见种分布格局及其对生物多样性计算的影响研究[D]. 海口: 海南大学, 2014.
- Zhou Wensong. Study on distribution patterns of rare species and their impact on the estimation of the species diversity in tropical evergreen monsoon dwarf forests in Tongguling, Hainan[D]. Haikou: Hainan University, 2014.
- [7] 陈正兴, 高德芳, 张伟, 等. 黄土丘陵沟壑区不同坡向撂荒地植物群落种群空间格局[J]. 应用生态学报, 2018, 29(6): 1846–1856.
- Chen Zhengxing, Gao Dexin, Zhang Wei, et al. Population spatial patterns of grassland plant communities in different slope aspects in the loess hilly area, China[J]. Chinese Journal of Applied Ecology, 2018, 29(6): 1846–1856.
- [8] Elith J, Leathwick J R. Species distribution models: ecological explanation and prediction across space and time[J]. Annual Review of Ecology, Evolution, and Systematics, 2009, 40: 677–697.
- [9] 郭彦龙, 赵泽芳, 乔慧捷, 等. 物种分布模型面临的挑战与发展趋势[J]. 地球科学进展, 2020, 35(12): 1292–1305.
- Guo Yanlong, Zhao Zefang, Qiao Huijie, et al. Challenges and development trend of species distribution model[J]. Advances in Earth Science, 2020, 35(12): 1292–1305.
- [10] 徐晓卉, 纪毓鹏, 薛莹, 等. 长山列岛邻近海域褐牙鲂资源丰度的时空分布及其与环境因子的关系[J]. 海洋学报, 2022, 44(2): 39–45.
- Xu Xiaohui, Ji Yupeng, Xue Ying, et al. Spatio-temporal distribution of *Paralichthys olivaceus* abundance and its relationship with environmental factors in the adjacent waters of Changshan Islands[J]. Haiyang Xuebao, 2022, 44(2): 39–45.
- [11] 崔晏华, 刘淑德, 张云雷, 等. 海州湾春季短蛸的栖息分布特征及其与环境因子的关系[J]. 应用生态学报, 2022, 33(6): 1686–1692.
- Cui Yanhua, Liu Shude, Zhang Yunlei, et al. Habitat characteristics of *Octopus ocellatus* and their relationship with environmental factors during spring in Haizhou Bay, China[J]. Chinese Journal of Applied Ecology, 2022, 33(6): 1686–1692.
- [12] 纪毓鹏, 李明坤, 韩东燕, 等. 山东南部近海脊腹褐虾时空分布及其与环境因子的关系[J]. 中国海洋大学学报(自然科学版), 2020, 50(7): 56–62.
- Ji Yupeng, Li Mingkun, Han Dongyan, et al. The relationship between spatiotemporal distribution of *Crangon affinis* and environmental factors in the southern coastal waters of Shandong[J]. Periodical of Ocean University of China, 2020, 50(7): 56–62.
- [13] Lomba A, Pellissier L, Randin C, et al. Overcoming the rare species modelling paradox: a novel hierarchical framework applied to an Iberian endemic plant[J]. Biological Conservation, 2010, 143(11): 2647–2657.
- [14] Breiner F T, Guisan A, Bergamini A, et al. Overcoming limitations of modelling rare species by using ensembles of small models[J]. Methods in Ecology and Evolution, 2015, 6(10): 1210–1218.
- [15] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. GB/T 12763.6–2007, 海洋调查规范 第6部分: 海洋生物调查[S]. 北京: 中国标准出版社, 2008.
- General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. GB/T 12763.6–2007, Specifications for oceanographic survey-Part 6: marine biological survey[S]. Beijing: Standards Press of China, 2008.
- [16] 张涛, 沃佳, 刘逸文, 等. 海州湾偶见鱼类的种类组成及时空分布[J]. 应用生态学报, 2023, 34(3): 787–795.
- Zhang Tao, Wo Jia, Liu Yiwen, et al. Species composition and spatial and temporal distribution of occasional fishes in Haizhou Bay, Shandong, China[J]. Chinese Journal of Applied Ecology, 2023, 34(3): 787–795.
- [17] 李敏, 纪毓鹏, 徐宾铨, 等. 黄河口及邻近水域小型鳀鲱鱼类数量分布及其与环境因子的关系[J]. 海洋学报, 2016, 38(10): 52–61.
- Li Min, Ji Yupeng, Xu Binduo, et al. Spatio-temporal distribution of small-sized fish species in Clupeidae and Engraulidae and its relationships with environmental factors in Huanghe River Estuary and its adjacent waters[J]. Haiyang Xuebao, 2016, 38(10): 52–61.
- [18] 张云雷, 孙霄, 刘晓慧, 等. 气候变化对黄海中南部斑鲷产卵场适宜性的影响[J]. 水产学报, 2022, 46(2): 215–223.
- Zhang Yunlei, Sun Xiao, Liu Xiaohui, et al. Impacts of climate changes on the habitat suitability of spawning ground for *Konosirus punctatus* in the central and southern Yellow Sea[J]. Journal of Fisheries of China, 2022, 46(2): 215–223.
- [19] 刘俊超, 贾明秀, 冯卫东, 等. 基于RF和GAM模型的南极磷虾资源分布与环境因子关系研究[J]. 中国海洋大学学报(自然科学版), 2021, 51(8): 20–29.
- Liu Junchao, Jia Mingxiu, Feng Weidong, et al. Spatial-temporal distribution of Antarctic Krill (*Euphausia superba*) resource and its association with environment factors revealed with RF and GAM models[J]. Periodical of Ocean University of China, 2021, 51(8): 20–29.
- [20] Guisan A, Edwards Jr T C, Hastie T. Generalized linear and generalized additive models in studies of species distributions: setting the scene[J]. Ecological Modelling, 2002, 157(2/3): 89–100.
- [21] 李丽霞, 郝艳晖, 周舒冬, 等. 广义加性模型及其应用[J]. 中国卫生统计, 2007, 24(3): 243–244.
- Li Lixia, Gao Yanhui, Zhou Shudong, et al. Generalized additive models and its application[J]. Chinese Journal of Health Statistics, 2007, 24(3): 243–244.
- [22] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5–32.

- [23] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, 26(3): 32–38.  
Fang Kuangnan, Wu Jianbin, Zhu Jianping, et al. A review of technologies on random forests[J]. *Statistics & Information Forum*, 2011, 26(3): 32–38.
- [24] 王奕森, 夏树涛. 集成学习之随机森林算法综述[J]. 信息通信技术, 2018, 12(1): 49–55.  
Wang Yisen, Xia Shutao. A survey of random forests algorithms[J]. *Information and Communications Technologies*, 2018, 12(1): 49–55.
- [25] 董师师, 黄哲学. 随机森林理论浅析[J]. 集成技术, 2013, 2(1): 1–7.  
Dong Shishi, Huang Zhexue. A brief theoretical overview of random forests[J]. *Journal of Integration Technology*, 2013, 2(1): 1–7.
- [26] Kabacoff R I. *R in Action: Data Analysis and Graphics with R*[M]. Greenwich: Manning Publications, 2011.
- [27] Wood S N. Stable and efficient multiple smoothing parameter estimation for generalized additive models[J]. *Journal of the American Statistical Association*, 2004, 99(467): 673–686.
- [28] Burnham K P, Anderson D R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*[M]. 2nd ed. New York: Springer, 2002.
- [29] Bradley A P. The use of the area under the ROC curve in the evaluation of machine learning algorithms[J]. *Pattern Recognition*, 1997, 30(7): 1145–1159.
- [30] Fielding A H, Bell J F. A review of methods for the assessment of prediction errors in conservation presence/absence models[J]. *Environmental Conservation*, 1997, 24(1): 38–49.
- [31] 栾静, 张崇良, 徐宾铎, 等. 海州湾双斑螯栖息分布特征与环境因子的关系[J]. 水产学报, 2018, 42(6): 889–901.  
Luan Jing, Zhang Chongliang, Xu Binduo, et al. Relationship between catch distribution of portunid crab (*Charybdis bimaculata*) and environmental factors based on three species distribution models in Haizhou Bay[J]. *Journal of Fisheries of China*, 2018, 42(6): 889–901.
- [32] 倪一卓, 程和琴, 江红, 等. 鱼类栖息地模拟的比较研究——以东海鲈鱼为例[J]. 水产科学, 2009, 28(12): 726–732.  
Ni Yizhuo, Cheng Heqin, Jiang Hong, et al. Fish habitat modeling in the East China Sea: a case of chub mackerel[J]. *Fisheries Science*, 2009, 28(12): 726–732.
- [33] Shono H. Is model selection using Akaike's information criterion appropriate for catch per unit effort standardization in large samples?[J]. *Fisheries Science*, 2005, 71(5): 978–986.
- [34] Swets J A. Measuring the accuracy of diagnostic systems[J]. *Science*, 1988, 240(4857): 1285–1293.
- [35] Li Min, Zhang Chongliang, Xu Binduo, et al. Evaluating the approaches of habitat suitability modelling for whitespotted conger (*Conger myriaster*)[J]. *Fisheries Research*, 2017, 195: 230–237.
- [36] Jacobucci R, Grimm K J. Machine learning and psychological research: the unexplored effect of measurement[J]. *Perspectives on Psychological Science*, 2020, 15(3): 809–816.
- [37] 袁传宓, 秦安聆. 我国近海鲚鱼生态习性及其产量变动状况[J]. 海洋科学, 1984, 8(5): 35–37.  
Yuan Chuanbi, Qin Anling. Ecological habits and distribution of *Coilia* along the Chinese coast and its changes of output[J]. *Marine Sciences*, 1984, 8(5): 35–37.
- [38] 徐敏. 盐碱度对3种鱼精子活力、受精率及孵化率的影响[D]. 上海: 上海海洋大学, 2015.  
Xu Min. Effects of salinity and alkalinity on sperm motility, fertilization rate and hatching rate of 3 fish species[D]. Shanghai: Shanghai Ocean University, 2015.
- [39] 成庆泰, 周才武. 山东鱼类志[M]. 济南: 山东科学技术出版社, 1997.  
Cheng Qingtai, Zhou Caiwu. *The Fishes of Shandong Province*[M]. Ji'nan: Shandong Science and Technology Press, 1997.
- [40] Stefansdottir L, Solmundsson J, Marteinsdottir G, et al. Groundfish species diversity and assemblage structure in Icelandic waters during recent years of warming[J]. *Fisheries Oceanography*, 2010, 19(1): 42–62.
- [41] 吴艳, 王洪峰, 穆立蓄. 物种分布模型的研究进展与展望[J]. 高师理科学刊, 2022, 42(5): 66–70.  
Wu Yan, Wang Hongfeng, Mu Liqiang. Research progress and prospect of species distribution models[J]. *Journal of Science of Teachers' College and University*, 2022, 42(5): 66–70.
- [42] 简盈, 张云雷, 宋业晖, 等. 基于GAM和GWR模型分析环境因子对鱼类分布的影响——以山东近海多鳞鱈为例[J]. 海洋学报, 2022, 44(7): 103–111.  
Jian Ying, Zhang Yunlei, Song Yehui, et al. Effect of environmental factors on fish distribution based on GAM and GWR model: a case study of *Sillago sihama* in the Shandong coastal waters[J]. *Haiyang Xuebao*, 2022, 44(7): 103–111.
- [43] Guo Yanlong, Li Xin, Zhao Zefang, et al. Prediction of the potential geographic distribution of the ectomycorrhizal mushroom *Tricholoma matsutake* under multiple climate change scenarios[J]. *Scientific Reports*, 2017, 7: 46221.
- [44] Stockwell D R B, Peterson A T. Effects of sample size on accuracy of species distribution models[J]. *Ecological Modelling*, 2002, 148(1): 1–13.
- [45] 尹洁, 王晶, 张崇良, 等. 利用two-stage GAM研究海州湾及其邻近海域小黄鱼鱼卵的时空分布特征[J]. 中国水产科学, 2019, 26(6): 1164–1174.  
Yin Jie, Wang Jing, Zhang Chongliang, et al. Spatial and temporal distribution characteristics of *Larimichthys polyactis* eggs in Haizhou Bay and adjacent regions based on two-stage GAM[J]. *Journal of Fishery Sciences of China*, 2019, 26(6): 1164–1174.
- [46] 张云雷, 徐宾铎, 张崇良, 等. 基于Tweedie-GAM模型研究海州湾小黄鱼资源丰度与栖息环境的关系[J]. 海洋学报, 2019, 41(12):

78–89.

Zhang Yunlei, Xu Binduo, Zhang Chongliang, et al. Relationship between the habitat factors and the abundance of small yellow croaker (*Larimichthys polyactis*) in Haizhou Bay based on the Tweedie-GAM model[J]. Haiyang Xuebao, 2019, 41(12): 78–89.

## Analysis of the influence of environmental factors on the distribution of occasional species in the Haizhou Bay based on species distribution model

Zhang Tao<sup>1,2</sup>, Zhao Tianya<sup>1,2</sup>, Luan Jing<sup>1,2</sup>, Zhang Yunlei<sup>1,2</sup>, Zhang Chongliang<sup>1,2</sup>

(1. Fisheries College, Ocean University of China, Qingdao 266003, China; 2. Field Observation and Research Station of Haizhou Bay Fishery Ecosystem, Ministry of Education, Qingdao 266003, China)

**Abstract:** Occasional species are vulnerable to external threats such as environmental changes and human activities and have important values in biodiversity conservation. However, due to their limited availability of data and associated difficulties in statistical analysis, there are few studies on the spatial distribution and their relationships with environmental factors. In this study, based on the fishery resource surveys in the Haizhou Bay conducted from 2013 to 2019, we analyzed the relationships between the distribution and environmental factors for three occasional species, *Coilia mystus*, *Odontamblyopus rubicundus* and *Erisphex pottii*, using generalized additive model (GAM) and random forest (RF) model. The models were compared according to their goodness of fit and the predictive performances were evaluated using cross-validation. The results showed that depth was the most significant factor affecting the distribution of *C. mystus* and *O. rubicundus* in spring and autumn, the sea bottom temperature was the most important environmental factor influencing the distribution of *E. pottii* in autumn. The distribution model of *C. mystus* had the highest deviance explanation, followed by *O. rubicundus*, and *E. pottii* had the lowest deviance explanation. The deviance explanation by the distribution models of *C. mystus*, *O. rubicundus* and *E. pottii* were all lower in spring than in autumn. The cross-validation showed that the area under the curve (AUC) of the three species ranged from 0.70 to 0.85, and only the AUC of *C. mystus* reached 0.9 in autumn; meanwhile, the AUC of the GAM prediction results were larger than those of the RF model, indicating that the prediction performance of the GAM was better than that of the RF model for the occasional species. This study would provide a reference for the selection of models for future studies of occasional species, and have guiding significance for the conservation of the occasional species.

**Key words:** Haizhou Bay; occasional species; generalized additive model (GAM); random forest (RF) model; cross-validation