

DOI: 10.12264/JFSC2025-0042

南太平洋长鳍金枪鱼栖息地分布影响因素

刘力文¹, 郑淳文¹, 李娅琳¹, 周想¹, 吴峰^{1, 2, 3}, 朱江峰^{1, 2, 3}, 周成^{1, 2, 3*}

1. 上海海洋大学海洋生物资源与管理学院, 上海 201306;
2. 上海海洋大学国家远洋渔业工程技术研究中心, 上海 201306;
3. 上海海洋大学, 农业农村部大洋渔业开发重点实验室, 上海 201306

摘要: 本研究基于 2020—2022 年南太平洋长鳍金枪鱼(*Thunnus alalunga*)延绳钓渔捞日志数据, 采用地理加权随机森林(geographically weighted random forest, GWRF)模型, 结合 Shapley 模型可解释(Shapley additive explanations, SHAP)技术, 构建了可解释的长鳍金枪鱼栖息地预测模型, 分析了关键环境因子对其分布的影响, 旨在为南太平洋长鳍金枪鱼栖息地研究及其可持续管理提供科学依据。研究结果表明, 在精度、准确率、召回率及受试者工作特征曲线下面积(area under the curve, AUC)等关键指标上, 地理加权随机森林模型相较于传统随机森林(random forest, RF)模型提高了 5%~10%。因子重要性分析 SHAP 贡献度分析表明, 海表温度、海表溶解氧浓度、50 m 深度温度及 50 m 深度溶解氧浓度是影响长鳍金枪鱼栖息地分布的关键环境因子; SHAP 解释性分析进一步揭示了适宜的栖息环境特征, 即海表温度和 50 m 深度温度处于 15~20 °C, 海表溶解氧浓度及 50 m 深度溶解氧浓度处于 240~260 mmol/m³的环境是最适宜长鳍金枪鱼栖息的。单一样本的 SHAP 值分解分析进一步验证了适宜的温度及充足的溶解氧是影响长鳍金枪鱼栖息地选择的关键因素。本研究为深入理解其栖息地的空间分布格局及环境驱动机制提供了新的视角。

关键词: 长鳍金枪鱼; 地理加权随机森林; 机器学习; 可解释模型

中图分类号: S931

文献标志码: A

文章编号: 1005-8737-(2025)08-1164-10

长鳍金枪鱼(*Thunnus alalunga*)是一种高经济价值的洄游性鱼类, 同时也是远洋延绳钓渔业的主要捕捞对象^[1]。然而, 随着捕捞压力的持续增加以及海洋环境的变化, 该物种的可持续管理面临诸多挑战^[2]。根据 2024 年南太平洋长鳍金枪鱼种群评估结果, 该种群目前尚未遭受过度捕捞, 但捕捞压力正逐年上升^[3]。为应对这一趋势, 未来可能引入基于配额的管理制度, 以确保资源的可持续利用。因此, 深入探究长鳍金枪鱼栖息地分布的关键驱动因素, 对于科学制定捕捞策略并推动渔业的可持续发展至关重要^[4]。

海洋环境的动态变化对长鳍金枪鱼的栖息地

分布具有重要影响。渔业研究通常依托渔捞日志、卫星遥感数据和海洋环境数据, 并结合栖息地模型来揭示其分布规律^[5-6]。栖息地模型通过分析历史数据中环境变量的作用机制, 构建预测框架, 以推测物种的潜在分布^[7]。Gavaris^[8]率先采用广义线性模型(generalized linear model, GLM)对单位捕捞努力渔获量(catch per unit effort, CPUE)数据进行标准化, 从而有效解决了因渔具差异导致的数据偏差。然而, CPUE 与环境因素之间的关系通常较为复杂, 而 GLM 的线性假设可能无法充分捕捉这种复杂性。为克服这一局限性, 广义加性模型(generalized additive model, GAM)通过引

收稿日期: 2025-03-17; 修订日期: 2025-05-22.

基金项目: 国家重点研发计划项目(2023YFD2401301, 2023YFD2401305); 农业农村部全球渔业资源调查、监测与评估项目(D-8025-24-5001).

作者简介: 刘力文(1996-), 男, 硕士研究生, 研究方向为金枪鱼渔业信息化. E-mail: gfdyjava@163.com

通信作者: 周成, 男, 副教授, 研究方向为金枪鱼渔业. E-mail: c-zhou@shou.edu.cn

入非线性平滑函数, 增强了模型对金枪鱼丰度与环境因素之间关系的刻画能力, 并能够识别关键驱动因素。Zagaglia 等^[9]和 Setiawati 等^[10]采用 GAM 分析了热带金枪鱼的栖息地分布特征, 研究表明, 该类鱼种的分布受热带辐合带影响, 并受到海表温度、海表高度异常、盐度和叶绿素浓度等环境变量的驱动。然而, 传统数理统计模型在处理复杂非线性关系时仍存在一定局限性, 特别是对于高度洄游性物种如长鳍金枪鱼, 难以精准刻画其种群动态的细微变化。相较于传统统计模型, 机器学习方法能够处理大规模、高维度、非线性数据, 尤其在探索环境因素(如温度、溶解氧、叶绿素浓度、洋流等)对长鳍金枪鱼栖息地的影响时, 表现出显著优势^[11-14]。其中随机森林模型, 因其稳定性和强大的预测能力而被广泛应用。Liu 等^[15]、侯娟等^[16]和宋利明等^[17]利用随机森林构建了长鳍金枪鱼渔场预测模型, 在长鳍金枪鱼渔场预测中表现出较高的预测精度和良好的鲁棒性。

机器学习模型在预测能力方面具有显著优势, 但其黑箱特性限制了模型的可解释性, 使得管理者难以理解预测结果的形成机制, 这在渔业资源管理实践中可能削弱其应用价值。为弥补这一局限性, 近年来, 研究者开始将机器学习模型与可解释性技术相结合, 以提升模型的透明度。Shapley 模型可解释技术(SHAP)作为一种广泛应用的解释性方法^[18], 可用于量化各环境变量对预测结果的贡献, 从而识别关键驱动因素^[19], 为渔业管理提供更具参考价值的科学数据。

1 材料方法

1.1 研究区域概况

本研究选取南太平洋部分海域作为研究区域, 研究范围覆盖赤道至南纬 40°、东经 160°至西经 120°。该区域不仅是长鳍金枪鱼的重要分布和渔业捕捞区, 同时也是多种热带、亚热带及温带鱼类的关键栖息地。研究区域内海洋环境复杂多样, 受到多种物理、化学及生物因子的综合影响, 包括海表温度、溶解氧浓度、叶绿素浓度及海面高度异常等。这些环境因子相互作用, 共同塑造了

长鳍金枪鱼的分布格局与栖息环境。

1.2 数据来源

1.2.1 渔业数据 本研究使用的渔业数据来源于上海海洋大学远洋渔业数据中心, 涵盖 2020—2022 年中国远洋温带延绳钓渔船的捕捞记录, 主要目标鱼种为长鳍金枪鱼。数据的时间分辨率为日尺度, 渔捞日志详细记录了捕捞位置(经纬度)、日期、渔获量(以尾数计)及相关作业信息。

1.2.2 环境数据 如表 1 所示, 本研究选取了不同水层(0 m、50 m、100 m、150 m、200 m、300 m)的多种海洋环境因子, 包括海水温度(thetao)、溶解氧浓度(o2)、海水盐度(so)、叶绿素 a 浓度(chl)、东向和北向海流速度(uv)、海面高度异常(ssha), 以及混合层厚度(mld)。所有数据的时间分辨率为逐日(每日 1 次), 空间分辨率为 0.25°×0.25°, 数据来源于哥白尼海洋环境监测服务(Copernicus marine environment monitoring service)。数据下载由 Python 实现, 环境因子与渔捞日志数据的匹配及后续处理则采用 R 语言(版本 4.3.1)。

表 1 用于模型构建的海洋环境因子

Tab. 1 Marine environmental factors applied in the model construction in the study

缩写 abbreviation	解释 explanation
mld	混合层厚度 ocean mixed layer thickness
ssha	海面高度异常 sea surface height above geoid
so	海水盐度 sea water salinity
thetao	海水温度 sea water potential temperature
chl	叶绿素 a 浓度 chlorophyll a concentration
o2	溶解氧浓度 dissolved oxygen concentration
uv	洋流流速 water velocity

1.3 数据预处理

计算 2020—2022 年的渔获量(尾数)和作业鱼钩数, 即单位捕捞努力量渔获量(catch per unit effort, CPUE)^[20]。具体计算公式如下。

$CPUE_{(i,j,d)} = Catch_{(i,j,d)} / Hooks_{(i,j,d)} \times 1000$ (1)
式中, i 和 j 表示作业渔船的经纬度位置, d 表示时间(天); Catch 表示捕捞个体数量; Hooks 表示钩子的总数量。

1.4 研究方法

1.4.1 模型方法 本研究采用随机森林(random

forest, RF)和地理加权随机森林(geographically weighted random forest, GWRF)模型,探讨长鳍金枪鱼渔获率与海洋环境因子之间的关系。随机森林是一种广泛应用的机器学习方法,因其卓越的非线性数据处理能力和高预测精度而受到认可。然而,该方法未能充分考虑环境因子的空间变异性。相比之下,地理加权随机森林在传统随机森林的基础上引入了空间加权机制^[21],使其能够更精确地刻画环境因子对渔获率的局部影响,从而提升模型的空间解释能力。本研究基于 R 语言进行模型构建与分析,主要依托 caret、SpatialML、spgwr 和 randomForest 等核心包。

1.4.2 五折交叉验证 为系统评估模型性能并筛选最优模型,本研究采用五折交叉验证方法。相比于更高折数的交叉验证,五折交叉验证在确保模型评估稳定性的同时,有效控制了计算成本。训练数据集被划分为 5 个分层子集,每次选取其中 4 个子集进行模型训练,并使用剩余 1 个子集进行验证,循环重复 5 次。通过这种方式,模型能够在不同数据子集上进行全面评估,减少因单次数据划分导致的偏差,提高结果的稳健性。在交叉验证过程中,本研究计算了准确率、平衡准确率、精确率、召回率和 F1 得分等多个性能指标,从多个维度对模型进行系统性评估。这种多角度分析不仅能够全面衡量模型的预测能力,还能确保最终选定的模型在不同数据分布下均具有良好的泛化性能。

1.4.3 模型评估 为评估模型的预测性能,本研究采用多个评价指标进行对比分析。采用 ROC 曲线(receiver operating characteristic curve)评估模型的区分能力。ROC 曲线通过计算不同判别阈值下的假阳性率(false positive rate, FPR)和真正例率(true positive rate, TPR),直观反映模型在不同阈值设定下的分类表现,曲线下面积(area under the curve, AUC)则作为综合评估指标, AUC 值越接近 1,模型区分能力越强。此外,利用混淆矩阵计算模型的准确率、平衡准确率、召回率、精确率和 F1 得分。这些指标能够从不同维度量化模型在各类别上的预测能力,进一步揭示其分类性能。具体计算公式如下:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2)$$

$$\text{BalanceAccuracy} = \frac{1}{2} \left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right) \times 100\% \quad (3)$$

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\% \quad (4)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (5)$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \quad (6)$$

在以上公式中,真正例(TP)是指实际值为正且预测值也为正的样本数;真负例(TN)是指实际值为负且预测值也为负的样本数。假正例(FP)表示实际值为负但预测值为正的样本,即将负例错误预测为正例;假负例(FN)则指实际值为正但预测值为负的样本,即将正例错误预测为负例。这些指标用于分类模型的评价,帮助衡量预测结果的准确性以及错误类型。

1.4.4 模型解释 为了提升模型的可解释性,引入了 SHAP 可解释方法。SHAP 基于合作博弈论原理,为每个特征在单次预测中的贡献分配重要性分数^[18,22],从而提供全局和局部层面的模型解释。全局 SHAP 分析用于揭示各特征在整个数据集中的相对重要性,而局部 SHAP 分析则展示特征对单次预测结果的具体影响。通过量化每个环境因子对模型预测结果的贡献,能够深入解析环境变量的作用,进而揭示驱动长鳍金枪鱼栖息地分布的关键因素。使用 R 语言中的 shapviz 和 fastshap 包计算并可视化 SHAP 值,以提升模型的透明度和可解释性。

2 结果与分析

2.1 模型的对比

通过对模型训练数据的验证,计算并绘制了 ROC 曲线,以评估模型的分类性能(图 1)。结果表明,地理加权随机森林模型(GWRF)的 AUC 值为 0.93,显著高于传统随机森林模型(RF)的 AUC 值(0.79)。此外,在准确率、召回率和平衡准确率等多个评估指标上, GWRF 的性能均优于 RF。GWRF 能够更精准地捕捉数据中的空间特征影响,在分

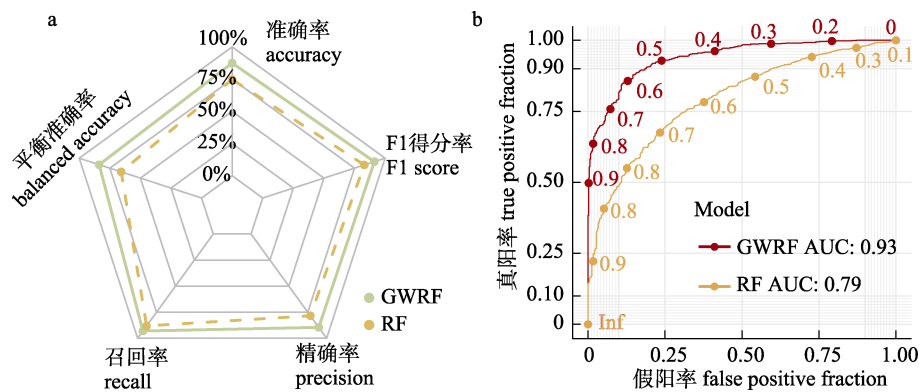


图 1 随机森林(RF)和地理加权随机森林(GWRF)模型的评价指标

a. RF 和 GWRF 模型的评价指标图; b. RF 和 GWRF 模型的受试者工作特征曲线图.

Fig. 1 Evaluation metrics of random forest (RF) and geographically weighted random forest (GWRF) models
a. Evaluation metrics of random forest and geographically weighted random forest model; b. Receiver operating characteristic curve of random forest and geographically weighted random forest model.

类性能和类别平衡方面表现出显著优势。相比之下, RF 模型未能充分考虑地理空间异质性, 限制了其预测能力, 使其分类性能略逊于 GWRF。

2.2 海洋环境因子的关键作用

如图 2 所示, 在前 10 个重要因子中, 不同水层的溶解氧和温度分别占据了前 7 位, 进一步凸

显了温度和溶解氧对长鳍金枪鱼栖息地分布的关键作用。此外, 150 m 水深的盐度、海面高度异常以及 300 m 水深的叶绿素浓度也被列为重要因子。这表明, 温度、溶解氧和叶绿素等多种环境因子共同作用, 塑造了长鳍金枪鱼栖息地的分布格局。

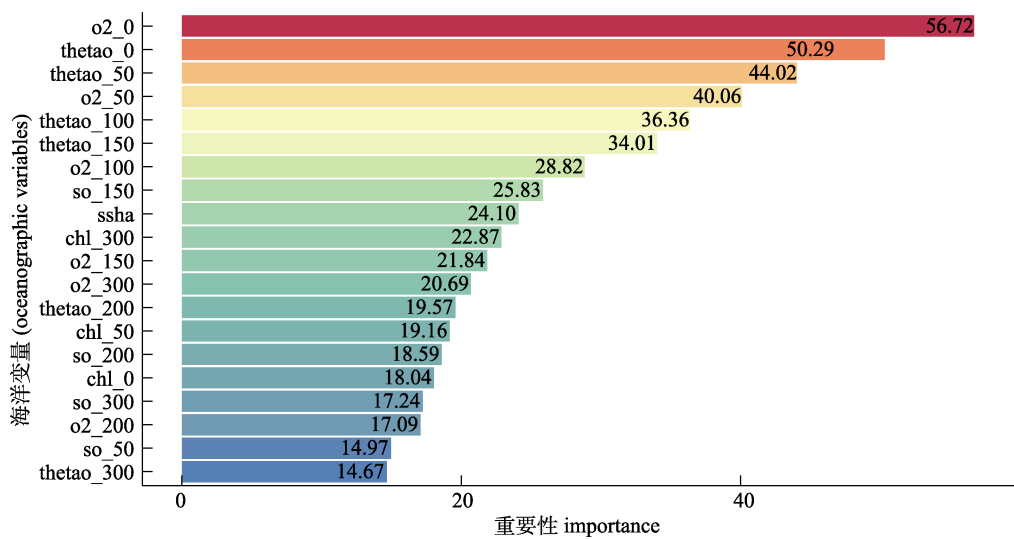


图 2 地理加权随机森林模型的海洋环境因子重要性柱状图

各环境因子缩写的解释见表 1.

Fig. 2 Bar chart of ocean environmental factor importance under the geographically weighted random forest model
The explanation for the environmental factor abbreviations are shown in Tab.1.

2.3 海洋环境因子的 SHAP 重要性分析

如图 3 所示, o2_0、o2_50、thetao_50 和 thetao_0 的点分布较为分散, 表明这 4 个特征对模型输出具有较大的贡献。其中, o2_0 和 o2_50 的

点分布呈现出一致的颜色趋势, 蓝色点集中在左侧, 红色点集中在右侧, 这表明较低的溶解氧值对预测结果产生负向影响, 而较高的溶解氧值则具有正向影响。相反, thetao_0、thetao_50 和

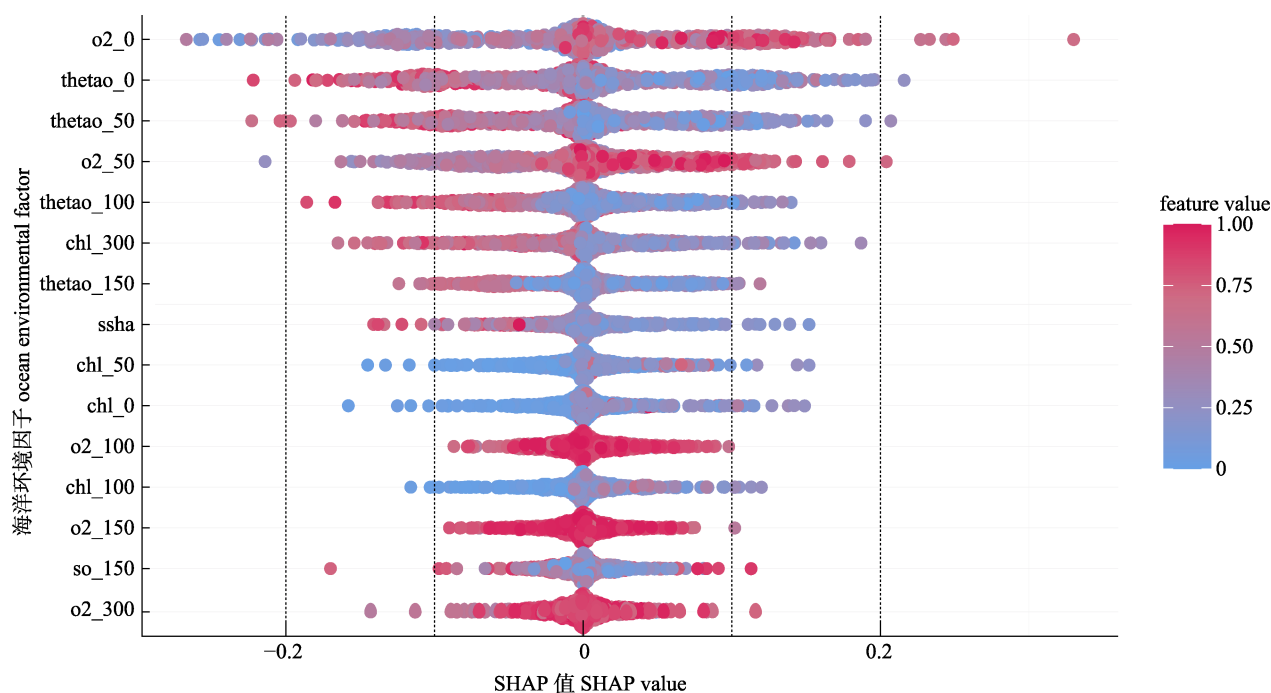


图 3 地理加权随机森林模型的海洋环境因子 SHAP 贡献值
各环境因子缩写的解释见表 1。

Fig. 3 SHAP contribution values of ocean environmental factors under the geographically weighted random forest (GWRF) model

The explanations for the environmental factor abbreviations are shown in Tab.1.

thetao_100 的颜色分布趋势与溶解氧相反, 蓝色点集中在右侧, 红色点集中在左侧, 表明较低的温度与适宜栖息地条件相关联。

2.4 海洋环境因子双变量依赖分析以及 2 维 SHAP 分析

使用 2 维 SHAP 分析结果如图 4 所示, 在图 4a 中 o2_0 与 o2_50 呈正相关, 随着 o2_0 的增加, o2_50 也随之增大, 同时 SHAP 值逐渐增大。当溶解氧浓度位于 240~260 mmol/m³ 区间时, SHAP 值超过 0.2, 表明该区间的溶解氧浓度对模型的正贡献较大。图 4b 显示, thetao_0 与 thetao_50 同样呈正相关, thetao_0 增加的同时, thetao_50 也随之增大, 但 SHAP 值逐渐减小。在温度 15℃ 至 20℃ 区间内, SHAP 值超过 0.2, 表明该温度范围对模型的正贡献较为显著。图 4c 表明, o2_0 与 thetao_0 之间存在负相关关系, 即温度升高时, 溶解氧浓度下降, 且 SHAP 值随之降低。类似地, 图 4d 揭示了 o2_50 与 thetao_50 之间的负相关关系, 温度上升时, 溶解氧浓度下降, SHAP 值也呈下降趋势。

当温度处于 15~20℃, 溶解氧浓度在

240~260 mmol/m³ 范围内时, SHAP 值均超过 0.2, 提示该温度和溶解氧浓度范围对模型预测具有显著的正向贡献。因此, 最适宜的栖息地温度约为 15~20℃, 最适宜的溶解氧浓度范围为 240~260 mmol/m³。

2.5 海洋环境变量对个体样本的影响

图 5 瀑布图和图 6 力图显示了单个预测实例中, 各特征(变量)对模型预测结果的贡献。它提供了直观的解释, 可视化每个特征如何推动预测结果, 力图通过显示特征对模型预测的整体影响, 强调各特征的交互作用。笔者随机抽取了单个预测实例并画图展示, 根据瀑布图和力图的结果累加所有特征的贡献, 得到 SHAP 的贡献值。其中深黄色代表正的贡献, 深红代表负的贡献。氧气、温度和叶绿素在特征贡献中的占比较大, 表明这些环境因子对长鳍金枪鱼栖息地的影响较为显著。此外, 这些因子之间存在一定的相互作用, 共同塑造了长鳍金枪鱼的分布格局。这些结果进一步验证了温度、溶解氧和叶绿素在确定长鳍金枪鱼适宜栖息地方面的重要性。

3 讨论

3.1 模型性能评估与空间异质性分析

随机森林模型在多项评估指标上均表现良

好, 包括准确率、平衡准确率、精确率、召回率及受试者工作特征曲线(ROC)分析, 表明其在区分正负样本方面具备较强能力, 进一步验证了其在长鳍金枪鱼分布预测中的可靠性。依托多决

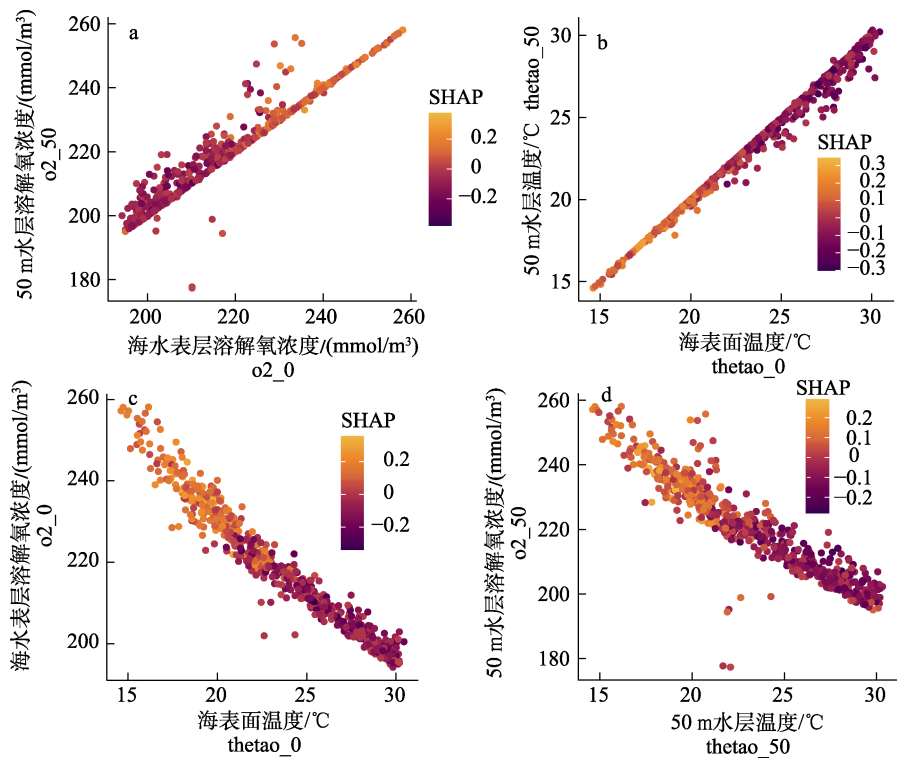


图 4 海洋环境因子双变量依赖关系图

- a. 海水表层和 50 m 水层溶解氧浓度的相互作用; b. 海水表层和 50 m 水层温度的相互作用;
c. 海水表层的溶解氧浓度和温度的相互作用; d. 50 m 水层的溶解氧浓度和温度的相互作用。

Fig. 4 Bivariate dependence plots of ocean environmental factors

- a. Interaction between sea surface and 50 m water layer dissolved oxygen concentration; b. Interaction between sea surface and 50 m water layer temperature; c. Interaction between dissolved oxygen concentration and temperature of sea surface; d. Interaction between dissolved oxygen concentration and temperature of 50 m water layer.

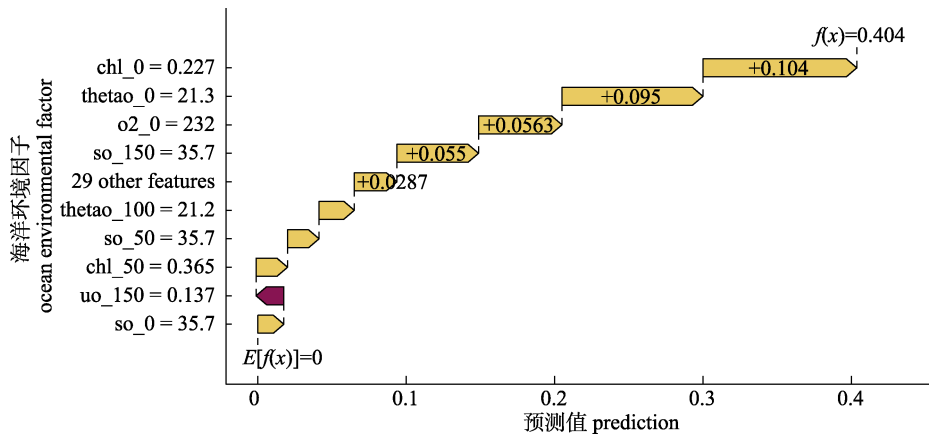


图 5 单一样本的 SHAP 瀑布图展示环境因子对预测结果的影响

各环境因子缩写的解释见表 1。

Fig. 5 The SHAP waterfall plot for a single sample showing the impact of environmental factors on the predicted result
The explanations for the environmental factor abbreviations are shown in Tab.1.

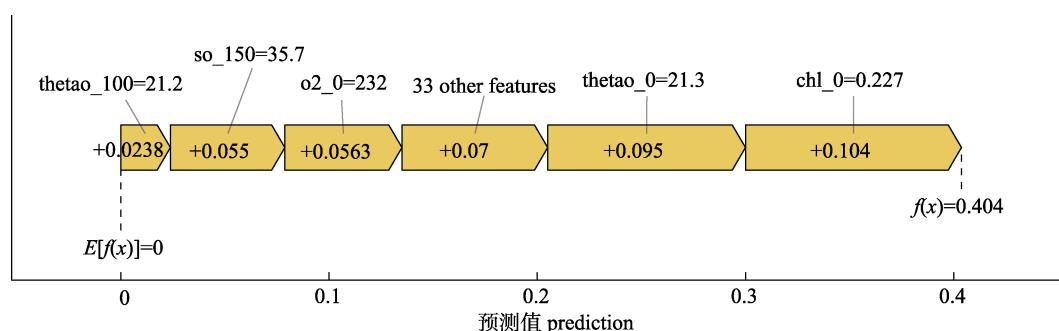


图 6 单一样本的 SHAP 力图展示环境变量对预测结果的影响

各环境因子缩写的解释见表 1。

Fig. 6 The SHAP force plot for a single sample showing the impact of environmental variables on the predicted result
The explanations for the environmental factor abbreviations are shown in Tab. 1.

策树投票机制, 随机森林模型展现出卓越的抗噪性和鲁棒性, 有效降低了对噪声和离群值的敏感性^[23-24], 这一特性在处理海洋环境数据固有的不确定性和噪声时尤为关键^[25]。此外, 本研究使用的大规模数据集覆盖广泛的时间和空间范围, 包括从海洋表面至 300 m 深度的多种环境因子, 随机森林模型能够挖掘这些变量之间的复杂关系, 从而提供精准的分布预测结果。

尽管随机森林模型在整体预测性能上表现优异, 其在捕捉空间异质性方面仍存在局限性。海洋环境因子的影响在不同区域可能存在显著差异, 忽略空间异质性可能会降低预测结果的适用性和准确性。为克服这一问题, 地理加权随机森林 (GWRF) 在传统随机森林的基础上引入了地理加权机制, 使模型能够根据不同区域的环境特征自适应地调整变量权重。这一改进不仅保留了随机森林强大的非线性建模能力与鲁棒性, 还能够精准刻画环境因子的空间异质性特征, 从而提高模型在不同空间尺度上的适应性^[26]。在长鳍金枪鱼分布预测中, GWRF 充分发挥了其空间敏感性优势, 相较于传统随机森林模型, 能够更精确地解析环境因子对鱼类分布的区域性影响。该方法为深入探索环境因子与物种分布之间的关系提供了更全面的分析框架, 也为基于空间异质性的渔场预报研究奠定了坚实的基础。

3.2 关键环境因子及其影响机制分析

在机器学习领域的快速发展中, 模型的复杂性往往使得预测过程呈现出“黑箱”特征。可解释

的机器学习方法在提高模型透明度和可理解性方面具有重要价值。本研究通过评估地理加权随机森林模型的特征重要性, 识别出影响长鳍金枪鱼栖息地分布的关键环境因子, 海表溶解氧(o2_0)和温度(thetao_0)是最关键的因素, 这些变量显著影响了长鳍金枪鱼的行为和生理特性^[27]。通过 SHAP 摘要图和双变量依赖图的分析, 进一步加深了对这些因子的理解。

在基于溶解氧的模型中, 最佳栖息地预测主要集中于海水表层^[28], 这一结果与本研究随机森林模型的预测一致(图 2)。溶解氧不仅影响长鳍金枪鱼的垂直分布, 还与其产卵区域密切相关^[29]。双变量依赖图分析表明, 当海表和 50 m 深度的溶解氧浓度在 240~260 mmol/m³ 范围内时, 是长鳍金枪鱼最为适宜的栖息环境。这一发现与 Lehodey 等^[30]的研究结果一致, 后者指出溶解氧浓度高于 4.5 mL/L 是西太平洋长鳍金枪鱼的最佳生存条件。相比其他金枪鱼类, 长鳍金枪鱼更偏好高溶解氧海域^[1]。在这些区域, 浮游生物的数量和多样性可能会增加, 氧气丰富的水体为其提供了理想的生态环境和生长条件。浮游生物是小型鱼类的主要食物来源, 鱼类会集中在浮游生物丰富的水域。溶解氧充足的区域因此成为小型鱼类聚集的场所, 吸引长鳍金枪鱼等掠食性鱼类, 使其成为重要的渔业资源区域。

Nugroho 等^[31]的研究表明, 15~26 °C 的海表温度可显著提高长鳍金枪鱼的捕获成功率, 温度直接影响其代谢功能和分布, 这与本研究得到的

15~20 °C 的最佳温度范围基本一致。此外, Boyce 等^[32]发现, 在南太平洋 21 °C 以下的温带海域, 长鳍金枪鱼的丰度较高。除此之外, CPUE 较高的海域通常具备相对适宜的海表温度, 通常在 18~21.47 °C 之间^[15,33]。这些研究表明, 适宜的温度海域是长鳍金枪鱼理想的栖息和繁衍地, 成为其栖息地的关键区域。

高盐度水域通常水团较为稳定, 并具有较高的生物适宜度, 可能吸引长鳍金枪鱼栖息其中^[34]。海面高度异常作为一种海洋物理现象, 通常指示着中尺度涡旋的存在。中尺度涡旋通过改变水体流动模式影响长鳍金枪鱼的洄游路径和分布范围。此外, 中尺度涡旋的动力搅拌作用能够促进初级生产力的聚集, 从而形成有利的觅食场所, 进一步吸引长鳍金枪鱼^[35-36]。叶绿素浓度作为初级生产力的重要指示因子, 间接反映了饵料生物的丰富程度, 对长鳍金枪鱼渔场的形成和分布具有重要影响^[37]。

本研究通过对森林随机模型引入地理加权机制, 有效刻画了环境变量对物种分布影响的空间差异性, 为进一步理解环境驱动机制提供了理论支持, 为提升渔场预报的空间适应性与生态合理性提供了新的方法框架, 对精细化渔业资源管理与可持续发展具有参考价值。

参考文献:

- [1] Nikolic N, Morandeau G, Hoarau L, et al. Review of albacore tuna, *Thunnus alalunga*, biology, fisheries and management[J]. Reviews in Fish Biology and Fisheries, 2017, 27(4): 775-810.
- [2] Pauly D, Christensen V, Guénette S, et al. Towards sustainability in world fisheries[J]. Nature, 2002, 418(6898): 689-695.
- [3] Teeares T, Castillo-Jordán C, Davies N, et al. Manila, Philippines 14–21 August 2024[J]. 2024.
- [4] Sund P N, Blackburn M, Williams F. Tunas and their environment in the Pacific Ocean: a review[J]. Oceanogr. Mar. Biol. Ann. Rev, 1981, 19: 443-512.
- [5] Mondal S, Wang Y C, Lee M G, et al. Ensemble three-dimensional habitat modeling of Indian Ocean immature albacore tuna (*Thunnus alalunga*) using remote sensing data[J]. Remote Sensing, 2022, 14(20): 5278.
- [6] Guan W J, Tian S Q, Wang X F, et al. A review of methods and model selection for standardizing CPUE[J]. Journal of Fishery Sciences of China, 2014, 21(4): 852-862. [官文江, 田思泉, 王学昉, 等. CPUE 标准化方法与模型选择的回顾与展望[J]. 中国水产科学, 2014, 21(4): 852-862.]
- [7] Elith J, H Graham C, P Anderson R, et al. Novel methods improve prediction of species' distributions from occurrence data[J]. Ecography, 2006, 29(2): 129-151.
- [8] Gavaris S. Use of a multiplicative model to estimate catch rate and effort from commercial data[J]. Canadian Journal of Fisheries and Aquatic Sciences, 1980, 37(12): 2272-2275.
- [9] Zagaglia C R, Lorenzetti J A, Stech J L. Remote sensing data and longline catches of yellowfin tuna (*Thunnus albacares*) in the equatorial Atlantic[J]. Remote Sensing of Environment, 2004, 93(1-2): 267-281.
- [10] Setiawati M D, Sambah A B, Miura F, et al. Characterization of bigeye tuna habitat in the Southern Waters off Java–Bali using remote sensing data[J]. Advances in Space Research, 2015, 55(2): 732-746.
- [11] Mohri M, Nishida T. Distribution of bigeye tuna and its relationship to the environmental conditions in the Indian Ocean based on the Japanese longline fisheries information [J]. Research and Development, 1999, 299(326pp): 319pp.
- [12] Chen I C, Lee P F, Tzeng W N. Distribution of albacore (*Thunnus alalunga*) in the Indian Ocean and its relation to environmental factors[J]. Fisheries Oceanography, 2005, 14(1): 71-80.
- [13] Zainuddin M, Saitoh K, Saitoh S I. Albacore (*Thunnus alalunga*) fishing ground in relation to oceanographic conditions in the western North Pacific Ocean using remotely sensed satellite data[J]. Fisheries Oceanography, 2008, 17(2): 61-73.
- [14] Zainuddin M, Safruddin S, Selamat M B, et al. Prediction of potential fishing zones for skipjack tuna during the northwest monsoon using remotely sensed satellite data[J]. Ilmu Kelautan: Indonesian Journal of Marine Sciences, 2017, 22(2): 59-66.
- [15] Liu L W, Wan R, Wu F, et al. An integrative machine learning approach to understanding South Pacific Ocean albacore tuna habitat features[J]. ICES Journal of Marine Science, 2025, 82(1): fsaf003.
- [16] Hou J, Zhou W F, Fan W, et al. Research on fishing grounds forecasting models of albacore tuna based on ensemble learning in South Pacific[J]. South China Fisheries Science, 2020, 16(5): 42-50. [侯娟, 周为峰, 樊伟, 等. 基于集成学习的南太平洋长鳍金枪鱼渔场预报模型研究[J]. 南方水产科学, 2020, 16(5): 42-50.]
- [17] Song L M, Ren S Y, Hong Y R, et al. Comparison on fishing

- ground forecast models of *Thunnus alalunga* in the tropical waters of Atlantic Ocean[J]. *Oceanologia et Limnologia Sinica*, 2022, 53(2): 496-504. [宋利明, 任士雨, 洪依然, 等. 大西洋热带海域长鳍金枪鱼渔场预报模型比较[J]. *海洋与湖沼*, 2022, 53(2): 496-504.]
- [18] M S, Lee S I. A unified approach to interpreting model predictions[C]//*Proceedings of the Neural Information Processing Systems.*, 2017
- [19] Zhang T J, Guo H, Song L M, et al. Evaluating the importance of vertical environmental variables for albacore fishing grounds in tropical Atlantic Ocean using machine learning and Shapley additive explanations (SHAP) approach[J]. *Fisheries Oceanography*, 2025, 34(1): e12701.
- [20] Feng Y J, Chen X J, Gao F, et al. Impacts of changing scale on Getis-Ord G_i^* hotspots of CPUE: A case study of the *Neon* flying squid (*Ommastrephes bartramii*) in the northwest Pacific Ocean[J]. *Acta Oceanologica Sinica*, 2018, 37(5): 67-76.
- [21] Santos F, Graw V, Bonilla S. A geographically weighted random forest approach for evaluate forest change drivers in the Northern Ecuadorian Amazon[J]. *PLoS One*, 2019, 14(12): e0226224.
- [22] Štrumbelj E, Kononenko I. Explaining prediction models and individual predictions with feature contributions[J]. *Knowledge and Information Systems*, 2014, 41(3): 647-665.
- [23] Belgiu M, Drăguț L. Random forest in remote sensing: A review of applications and future directions[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 114: 24-31.
- [24] Biau G, Scornet E. A random forest guided tour[J]. *TEST*, 2016, 25(2): 197-227.
- [25] Rubbens P, Brodie S, Cordier T, et al. Machine learning in marine ecology: An overview of techniques and applications[J]. *ICES Journal of Marine Science*, 2023, 80(7): 1829-1853.
- [26] Georganos S, Kalogirou S. A forest of forests: A spatially weighted and computationally efficient formulation of geographical random forests[J]. *ISPRS International Journal of Geo-Information*, 2022, 11(9): 471.
- [27] Chang Keyang, Chen C S, Wang Huiyu, et al. The Antarctic Oscillation index as an environmental parameter for predicting catches of the Argentine shortfin squid (*Illex argentinus*) (Cephalopoda: Ommastrephidae) in southwest Atlantic waters[J]. *Fishery Bulletin*, 2015, 113(2): 202-212.
- [28] Su S Y, Mao Q H, Li Y D, et al. Deep learning-based fishing ground prediction for albacore and yellowfin tuna in the Western and Central Pacific Ocean[J]. *Fisheries Research*, 2024, 278: 107103.
- [29] Block B A, Keen J E, Castillo B, et al. Environmental preferences of yellowfin tuna (*Thunnus albacares*) at the northern extent of its range[J]. *Marine Biology*, 1997, 130(1): 119-132.
- [30] Lehodey P, Senina I, Nicol S, et al. Modelling the impact of climate change on South Pacific albacore tuna[J]. *Deep Sea Research Part II: Topical Studies in Oceanography*, 2015, 113: 246-259.
- [31] Nugroho S C, Setiawan R Y, Setiawati M D, et al. Estimation of albacore tuna potential fishing grounds in the southeastern Indian ocean[J]. *IEEE Access*, 2023, 11: 1141-1147.
- [32] Boyce D G, Tittensor D P, Worm B. Effects of temperature on global patterns of tuna and billfish richness[J]. *Marine Ecology Progress Series*, 2008, 355: 267-276.
- [33] Zainuddin M, Kiyofuji H, Saitoh K, et al. Using multi-sensor satellite remote sensing and catch data to detect ocean hot spots for albacore (*Thunnus alalunga*) in the northwestern North Pacific[J]. *Deep Sea Research Part II: Topical Studies in Oceanography*, 2006, 53(3-4): 419-431.
- [34] Shcherbina A Y, D'Asaro E A, Riser S C, et al. Variability and interleaving of upper-ocean water masses surrounding the north Atlantic salinity maximum[J]. *Oceanography*, 2015, 28(1, Special Issue on the Salinity Processes in the Upper-ocean Regional Study): 106-113.
- [35] Tew Kai E, Marsac F. Influence of mesoscale eddies on spatial structuring of top predators' communities in the Mozambique Channel[J]. *Progress in Oceanography*, 2010, 86(1-2): 214-223.
- [36] Zhou C, He P G, Xu L X, et al. The effects of mesoscale oceanographic structures and ambient conditions on the catch of albacore tuna in the South Pacific longline fishery[J]. *Fisheries Oceanography*, 2020, 29(3): 238-251.
- [37] Yan M, Zhang H, Fan W, et al. Spatial-temporal CPUE profiles of the albacore tuna (*Thunnus alalunga*) and their relations to marine environmental factors in the South Pacific Ocean[J]. *Chinese Journal of Ecology*, 2015, 34(11): 3191-3197. [闫敏, 张衡, 樊伟, 等. 南太平洋长鳍金枪鱼渔场 CPUE 时空分布及其与关键海洋环境因子的关系[J]. *生态学杂志*, 2015, 34(11): 3191-3197.]

Factors influencing the habitat distribution of albacore tuna in the South Pacific

LIU Liwen¹, ZHENG Chunwen¹, LI Yalin¹, ZHOU Xiang¹, WU Feng^{1,2,3}, ZHU Jiangfeng^{1,2,3}, ZHOU Cheng^{1,2,3*}

1. College of Marine Living Resource Sciences and Management, Shanghai Ocean University, Shanghai 201306, China;
2. National Engineering Research Center for Oceanic Fisheries, Shanghai Ocean University, Shanghai 201306, China;
3. Key Laboratory of Exploitation of Ocean Fisheries Resources, Ministry of Agriculture and Rural Affairs; Shanghai Ocean University, Shanghai 201306. China

Abstract: This study is based on logbook data from longline fisheries targeting albacore tuna (*Thunnus alalunga*) in the South Pacific from 2020 to 2022. A geographically weighted random forest (GWRF) model was applied and Shapley additive explanations (SHAP) were incorporated to develop an interpretable habitat prediction model and analyze the influence of key environmental factors on albacore tuna distribution. The study aims to provide a scientific basis for habitat research and the sustainable management of albacore tuna in the South Pacific. The results indicate that compared to the traditional random forest (RF) model, the GWRF model improves key performance metrics, including precision, accuracy, recall, and the area under the receiver operating characteristic curve (AUC), by 5%–10%. Feature importance and SHAP contribution analyses identified sea surface temperature, sea surface dissolved oxygen concentration, temperature at 50 m depth, and dissolved oxygen concentration at 50 m depth as the key environmental factors influencing albacore tuna habitat distribution. SHAP interpretability analysis further revealed the optimal habitat conditions, indicating that the most suitable habitats were located in areas where sea surface temperature and temperature at 50 m depth ranged from 15 °C to 20 °C while sea surface dissolved oxygen concentration and dissolved oxygen concentration at 50 m depth ranged from 240 to 260 mmol/m³. Individual prediction SHAP value decomposition further confirmed that suitable temperature and sufficient dissolved oxygen are the key factors influencing albacore tuna habitat selection. These findings provide new insights into the spatial distribution patterns of albacore tuna habitats and the underlying environmental driving mechanisms.

Key words: *Thunnus alalunga*; geographically weighted random forest; machine learning; interpretable model

Corresponding author: ZHOU Cheng. E-mail: c-zhou@shou.edu.cn