

DOI: 10.12264/JFSC2024-0364

金枪鱼延绳钓渔业中 CPUE 的不同计算方法对渔场预测精度的影响

宋利明^{1,2}, 许回¹, 齐雨琨¹, 卢灿璜³, 申丰源⁴, 丁鹏¹, 李玉伟^{1,2*}, 沈介然⁵

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. 大洋世家(浙江)股份公司, 浙江 舟山 316014;
4. 中国水产科学研究院东海水产研究所, 上海 200090;
5. 深圳市联成远洋渔业有限公司, 广东 深圳 518035

摘要: 为提高延绳钓渔业的渔场预测精度, 本研究依据密克罗尼西亚海域延绳钓渔船的渔捞日志, 以及船舶监测系统 (vessel monitor system, VMS) 得到的渔业数据, 计算得出黄鳍金枪鱼 (*Thunnus albacares*) 的两种单位捕捞努力量渔获量 (catch per unit effort, CPUE)。基于双向长短期记忆网络 (Bi-directional long-short-term memory, BiLSTM) 模型, 将两种 CPUE 分别作为模型输入数据, 结合空间要素和海洋遥感环境数据, 建立黄鳍金枪鱼的渔场预报模型, 并评价不同的 CPUE 输入对渔场预测精度的影响; 通过 SHAP (SHapley Additive exPlanations) 分析, 得出对黄鳍金枪鱼渔场预测的重要变量; 通过不同 CPUE 的地理空间分布得出渔场的时空分布特征。结果表明, BiLSTM 模型可用于预测黄鳍金枪鱼渔场, 具有良好的预测效果; 基于 VMS 计算的 CPUE 比基于渔捞日志计算的 CPUE 建立的 BiLSTM 渔场预测模型有更高的精度。影响黄鳍金枪鱼渔场预测的重要变量是: 叶绿素浓度、200 m 水层的溶解氧浓度、300 m 水层的温度、200 m 水层的温度和经度等。基于渔捞日志的渔场南北分散, 而基于 VMS 的渔场主要集中在南部, 由温度和溶解氧浓度的垂直剖面图推测基于 VMS 得出的渔场更合理; 密克罗尼西亚海域黄鳍金枪鱼渔场在一二季度分布较为密集。本研究证明基于 VMS 可计算延绳钓渔业的 CPUE 并用于渔场预测; 基于本研究的结果, 当以较小空间尺度研究金枪鱼延绳钓渔场时, 建议按照 VMS 的方式逐条记录渔获位置, 以网格化计算 CPUE, 从而提高渔场预测精度。

关键词: 延绳钓; 黄鳍金枪鱼; BiLSTM; CPUE; 渔场预测

中图分类号: S931

文献标志码: A

文章编号: 1005-8737-(2025)05-0659-16

黄鳍金枪鱼 (*Thunnus albacares*) 是世界上重要的经济鱼种, 是金枪鱼延绳钓的主要捕捞对象之一^[1]。位于中西太平洋的密克罗尼西亚联邦专属经济区面积约 2780 km², 是世界上最丰富的金枪鱼渔场之一^[2]。黄鳍金枪鱼是密克罗尼西亚海域延绳钓渔业的主要捕捞鱼种, 中国和日本是在该海域开展金枪鱼延绳钓作业的主要国家。目前关于密克罗尼西亚海域金枪鱼的相关研究集中在

金枪鱼围网渔业管理策略^[3]、相关目标鱼种早期生态学^[4]等, 关于黄鳍金枪鱼延绳钓渔业及渔场的研究尚为空白, 因此有必要开展渔场预测等相关研究以促进了解该海域的黄鳍金枪鱼的资源现状, 为在此海域作业的中国延绳钓渔船提高金枪鱼产量提供参考。

单位捕捞努力量渔获量 (catch per unit effort, CPUE) 通常被视为渔业资源丰度指标而用于渔场

收稿日期: 2024-11-29; **修订日期:** 2025-04-09.

基金项目: 国家重点研发计划项目 (2023YFD2401301); 中水集团远洋股份有限公司科研项目“低温延绳钓渔船作业模式创新研究” (COFC-C-F-2024-004)

作者简介: 宋利明 (1968–), 男, 教授, 研究方向为捕捞学. E-mail: lmsong@shou.edu.cn

通信作者: 李玉伟, 讲师, 主要研究方向为渔具水动力学和金枪鱼栖息地等研究. E-mail: ywli@shou.edu.cn

预测^[5-8]。CPUE 的计算方法在不同的渔业类型中差异显著,这影响了渔业资源评估的准确性。例如田思泉等^[9]基于南大西洋鱿钓渔业的渔捞日志比较了 CPUE 的不同计算方法对标准化结果的影响,计算方法的差异在于是否划分小渔区。其研究强调应充分考虑不同 CPUE 计算方法带来的不确定性,以提高模型结果的可靠性。Lauridsen 等^[10]在研究湖泊刺网渔业中比较了两种不同标准下(区别在于是否综合使用底栖刺网数据和中层刺网数据)CPUE 计算方法的差异,研究表明包含中层刺网的数据更能反映深湖渔业资源分布的全貌。但延绳钓渔业 CPUE 计算方法的比较研究目前还很少。延绳钓渔业数据以非独立调查数据居多,包括渔业观察员数据、渔捞日志数据以及船舶配置的船舶监测系统(vessel monitor system, VMS)数据等,也被称为商业性渔业数据。尽管延绳钓渔业中 CPUE 的结果通常以每千钩的渔获尾数呈现,但由于渔业数据记录及统计方式的差异,一定时空范围内的渔获量及有效努力量的结果可能存在较大差异。围网等渔业通常以网次表示努力量^[11],而延绳钓渔业中通常以钩钩数量表征努力量^[12]。与其他捕捞金枪鱼的作业方式相比,商业金枪鱼延绳钓有持续时间长和作业范围大的特性,单次作业航程可达 60 海里,可包含多个精细网格^[13],这使得钩钩数量的网格化统计意义重大。不同计算方法可导致相同单元网格内尤其是较小空间尺度的单元网格的捕捞量和努力量不同,从而导致 CPUE 的估值差异,进而影响金枪鱼渔场预报的精度,因此有必要评估 CPUE 的不同计算方法带来的误差影响。延绳钓渔业数据来源有多种,例如渔捞日志每次主要记录作业开始的位置及渔获总量^[14],观察员数据记录篮数、下钩起始时间、作业开始和结束位置及渔获信息等^[15-16],VMS 监测数据记录了每尾渔获的位置,可判断渔船航迹^[17]。由于金枪鱼有集群的特性,鱼类密度分布并不均匀,这意味着渔获量在捕捞航程中非均匀分配。记录各个渔获信息的 VMS 为估算渔获位置之间的有效努力量提供了可能性,例如 Xu 等^[13]基于 VMS 数据推算延绳钓渔船的作业轨迹,按照航程的比例估算在单位网格内的钩钩数量,该捕捞努

力量的估算方法为延绳钓 CPUE 的计算提供了新的方法,但因缺乏与常规计算方法的对照与误差分析,是否为 CPUE 的理想计算方法需要进一步比较研究。

金枪鱼的渔场预报模型及方法有多种,人工神经网络(artificial neural network, ANN)模型^[18]、综合栖息地指数模型^[19]、提升回归树^[20]、集成模型^[21]等在金枪鱼渔场预测中均有较高的准确率。但上述模型也存在诸如难以理解具体的生物学机制、对输入数据质量的依赖性较高、不善于处理高维数据等缺点。集成模型的复杂性和对大规模标注数据的需求可能导致在数据不充分或偏差较大的情况下表现不佳。近年来,克服传统递归神经网络(recurrent neural network, RNN)中因梯度消失或爆炸导致的长期依赖问题的长短期记忆(long-short-term memory, LSTM)神经网络模型在金枪鱼渔场预报中得到了应用并表现出较好的性能^[22-23]。双向长短期记忆(bi-directional long-short-term memory, BiLSTM)神经网络模型是 LSTM 模型的扩展版本,将输入数据同时传递给正向处理序列和反向处理序列的两个 LSTM 网络,这种双向结构提高了模型性能^[24]并在金枪鱼渔场预测中表现出较好的性能^[25]。本文基于 BiLSTM 模型,将不同计算方法得到的黄鳍金枪鱼 CPUE 分别作为输入,结合空间及海洋遥感环境要素,建立黄鳍金枪鱼渔场预报模型并评估预测精度,并分析精度较高的模型中影响渔场预测的重要变量,为延绳钓渔业 CPUE 的计算以及黄鳍金枪鱼的渔场预测研究提供参考。

1 材料与方法

1.1 渔业数据与环境数据

1.1.1 渔业数据来源与特征 黄鳍金枪鱼渔业数据来自在密克罗尼西亚海域作业的中国深圳市联成远洋渔业有限公司的三艘延绳钓渔船的渔捞日志和 VMS,时间从 2019 年 1 月至 2022 年 12 月,作业海域范围为 $0^{\circ}15'N\sim 11^{\circ}14'N$, $143^{\circ}30'E\sim 163^{\circ}46'E$ 。原始数据的精度为 0.0001° ,将渔业数据的空间分辨率定义为 0.25° ,时间分辨率定义为

天。渔捞日志记录着延绳钓渔船每次作业的始位置、渔获量以及渔获的体重等信息。VMS 记录的信息有: 延绳钓渔船船名、每次作业投放的钓钩数量、捕获的鱼种、各鱼种的渔获时间与渔获位置(经纬度)等。延绳钓作业期间主要捕捞对象

为黄鳍金枪鱼和大眼金枪鱼(*Thunnus obesus*), 其中黄鳍金枪鱼是优势种, 在作业期间为主要渔获(图 1)。各渔船数据记录详尽, 包含高分辨率的捕捞活动, 涵盖密克罗尼西亚的大部分海域, 作业模式能够代表该海域延绳钓渔业的主要特征。

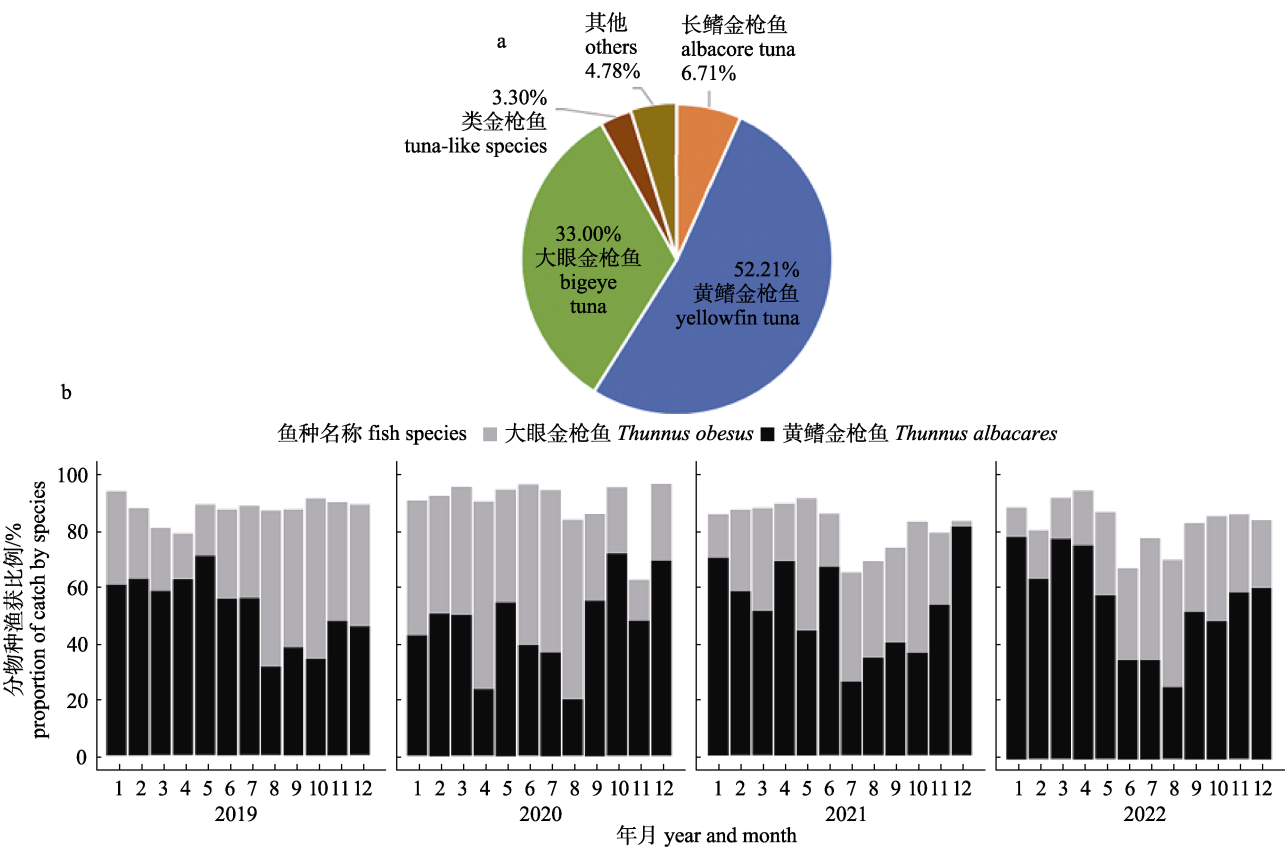


图 1 延绳钓作业期间渔获概况
a. 各类渔获比例; b. 主要捕捞物种在各月的比例。
Fig. 1 Overview of catch composition during longline fishery operations
a. Proportion of catch by taxa; b. Monthly proportion of main targeted species

1.1.2 延绳钓渔船作业参数 在捕捞黄鳍金枪鱼期间, 延绳钓作业参数如下: 平均投绳时间约 7 h, 投绳时平均船速为 4.4 m/s。投绳机平均出绳速度为 5.5 m/s, 投放相邻两钓钩的平均时间间隔为 6.0 s。起绳的平均时间约为 10.5 h, 平均船速约为 5.7 n mile/h。浮子绳长约 25 m, 支线长约 21 m, 每筐的干线长度约 860~924 m, 每筐的钓钩数量为 25~28 个。每次作业投放的钓钩数量为 2500~3300 个。

1.1.3 环境数据 由于各水层的盐度范围差值不大, 而影响中上层鱼类的浮游生物量往往和表层

叶绿素浓度关系密切^[26], 因此仅取海表面的盐度和叶绿素浓度。依据悬链线公式^[27], 估算延绳钓钓钩的最大沉降深度约 291.31 m。由于延绳钓在海流等作用下会变浅^[16], 选取能完全覆盖渔获深度的 0~300 m 水层的温度及溶解氧浓度作为候选变量, 每隔 50 m 取一次。环境数据为卫星遥感数据, 从 Copernicus Marine Environment Monitoring (CMEMS, 网址: <http://marine.copernicus.eu>)下载, 空间分辨率为 0.25°×0.25°, 时间分辨率为天。按照时间与地点, 将环境数据匹配到渔业数据中, 数据匹配在 R 语言中实现。

1.2 CPUE 的不同计算方法

1.2.1 基于渔捞日志计算 CPUE 基于渔捞日志的 CPUE 按照公式(1)计算, 对应的渔业数据与环境数据的集合记为方法 1 (Method 1)。

$$CPUE = \frac{\sum_{m=1}^n C_{i,j,k}}{\sum_{m=1}^n H_{i,j,k}} * 1000 \quad (1)$$

式中, $\sum_{m=1}^n C_{i,j,k}$ 表示第 i 年 j 月网格 k ($0.25^\circ \times 0.25^\circ$)

内的渔获总尾数, $\sum_{m=1}^n H_{i,j,k}$ 表示第 i 年 j 月网格 k

内的钩钩总数, CPUE 的单位为尾/千钩。

1.2.2 基于 VMS 计算 CPUE VMS 记录各尾渔获的时间与位置等信息, 由于延绳钓渔船起绳过程中经常跨天进行, 在计算每次作业的渔获量和努力量时先将跨天作业的日期统计为占据时间较多的那天。以渔船某天的起绳作业为例, 延绳钓的渔获位置及航迹如图 2 所示。起绳作业的总航程用式(2)计算。假设 VMS 中每两尾渔获间的航迹为直线, 则由各渔获点(图 2 b~q 的黑点)可计算出已知航程。总航程与已知航程之差为剩余航程(ab 与 qs 之和)。假设 ab 和 qs 航程相等, 且 abc 方向相同, qrs 方向也相同, 据此可以推断出起绳的起点 a 和终点 s 的位置。接着标记出渔船航迹与网格线的交点(c, i, m, r)。依据渔船在各网格内的航程占据总航程的比例计算努力量。例如网格 A 中的努力量用式(3)计算, 网格 E 中的努力量用式(4)计算。然后统计一定时间尺度和空间网格($0.25^\circ \times 0.25^\circ$)的渔获量和努力量, 计算 CPUE, 见式(5)^[12]。基于 VMS 对应的数据集记为方法 2, 相关计算在 R 中完成。

$$S_T = \left(\frac{N_H}{n_H} + 1 + N_H + 1 \right) * t_d * V_d \quad (2)$$

式中, N_H 表示每次作业时的放钩总数, $\frac{N_H}{n_H} + 1$ 表示每次作业时的浮子数, n_H 表示每筐的钩钩数量, t_d 表示相邻两钩钩的投钩时间间隔, $\left(\frac{N_H}{n_H} + 1 + N_H + 1 \right) * t_d$ 表示每次作业时的投绳总时间, V_d 表示起绳期间

的平均船速。

$$H_A = \frac{S_{ab} + S_{bc}}{S_T} * N_H \quad (3)$$

$$H_E = \frac{S_{rs}}{S_T} * N_H \quad (4)$$

$$CPUE_{x,y} = \frac{C_{x,y}}{H_{x,y}} * 1000 \quad (5)$$

式中, $CPUE_{x,y}$ 表示网格中心(x, y)的 CPUE; $C_{x,y}$ 指点(x, y)分别向四周扩展 0.125° (即以 $0.25^\circ \times 0.25^\circ$ 网格为中心)的网格内各渔船每天的总渔获尾数, $H_{x,y}$ 指网格中心(x, y)中每艘渔船每天的总投钩数。

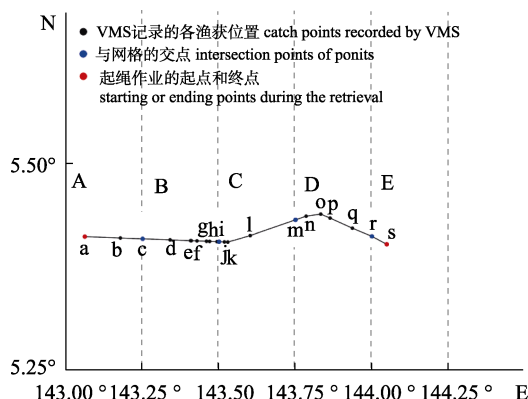


图 2 根据 VMS 估算延绳钓努力量的示意图

A-E 表示不同的网格, a-s 表示延绳钓起绳开始至结束的重要坐标点, 其中红点 a 和 s 分别表示起绳作业的开始和结束位置, 黑点表示 VMS 记录的各渔获位置, 蓝点表示相邻两点的连线与网格的交点。

Fig. 2 Schematic diagram of estimating effort in longline fishing based on VMS

A-E represent different grid cells, while a-s denote key coordinate points from the start to the end of the longline retrieval operation. The red points, a and s, denote the starting and ending positions of the retrieval operation, respectively. Black points represent the catch locations recorded by VMS, and blue points indicate the intersections between the lines connecting adjacent points and the grid boundaries.

1.2.3 CPUE 划区及渔场定义 将 CPUE 数据按照三分位点进行划分, 其中, 第 1 分位点(Q1, 33.3%)和第 2 分位点(Q2, 66.6%)分别用于区分 CPUE 的低值($\leq Q1$)、中值(Q1 与 Q2 之间)和高值($> Q2$)。将 CPUE 低值区定义为非渔场区, CPUE 中值及高值区定义为渔场区, CPUE 高值区为渔场密集区。

1.3 数据预处理

1.3.1 数据归一化 在深度学习中, 在建模前对数据进行归一化处理有助于稳定和加快模型的训练过程。归一化通过方程式(6)将数据值转化在[0,

1]之间。在应用 Bi-LSTM 模型后, 对预测结果进行反归一化处理, 确保预测数据与实际范围一致, 使得有关分析符合实际。

$$X_j = \frac{x_j - x_{\min}}{x_{\max} - x_{\min}} \quad (6)$$

式中, X_j 表示某变量归一化后的数值, x_j 表示某变量第 j 条数据初始值, $x_{\max}$ 表示某变量的最大值, $x_{\min}$ 表示某变量的最小值。

1.3.2 特征变量选择 深度学习在建模前进行特征变量选择有利于提升模型性能^[28]。特征选择过程通过从数据集中删除不重要的特征来最小化数据复杂性以降低检测入侵时的误报率^[29]。本文选择 LASSO 回归进行特征选择。LASSO 回归通过将某些变量的回归系数缩小至零, 从而有效地排除不利于建模的变量, 在特征选择上性能较好^[30]。在深度学习中, LASSO 回归特别适用于高维数据场景, 能有效避免多重共线性的问题并减少过拟合, 相较其他变量选择方法具有优越性。例如主成分分析主要用于降维, 但可能导致信息丢失; 而随机森林在高维数据下计算成本较高, 且可能导致模型复杂度增加^[31]。LASSO 的估计通过公式(7)计算。

$$\hat{\beta}^{\text{lasso}} = \arg_{\beta} \min \left\{ \frac{1}{2} \sum_{i=1}^N \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (7)$$

式中, $\hat{\beta}^{\text{lasso}}$ 表示 LASSO 回归中的估计系数; y_i 表示第 i 个样本中因变量的观测值; β_0 表示回归模型中的截距项; x_{ij} 表示第 i 个样本中的第 j 个预测变量; β_j 表示第 j 个预测变量的回归系数; λ 表示调整参数, 用于控制系数的收缩量。当 λ 增大时, 系数受到更大的惩罚并趋向于零; $\sum_{j=1}^p |\beta_j|$ 表示 L1 惩罚项, 即回归系数绝对值的总和。

完成特征选择后, 依据 VIF 数值对各个变量进行多重共线性诊断以排除时空环境因子之间可能存在的高相关性。当所有解释变量的 VIF 值均小于 5 时, 认为排除了多重共线性^[32]。

1.4 Bi-LSTM 渔场预测模型

1.4.1 Bi-LSTM 建模 BiLSTM 是由 LSTM 发展而来。LSTM 是特殊的 RNN, 通过加入“门”(gate)数据结构来提高在同一层中前一个细胞(cell)信息传递给下一个细胞的效率, 有效缓解梯度消失和梯度爆炸的问题^[33]。LSTM 主体由遗忘门、输入门和输出门构成, 模型内部结构如图 3 所示。各门的计算公式参考文献^[13]。

BiLSTM 是一种改进的循环神经网络(RNN)结构, 能够从时间序列数据中捕捉到长期和短期的依赖关系^[34], 结构见图 4。BiLSTM 具有与 LSTM 相同的存储单元架构, 但通过在输入序列的前向

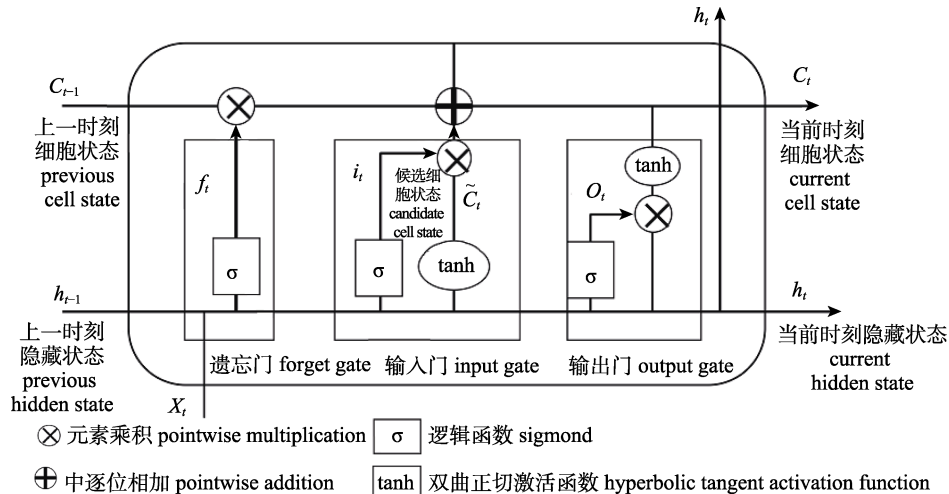


图 3 LSTM 模型结构

f_t 、 i_t 、 O_t 分别表示当前时刻遗忘门、输入门和输出门的信息。

Fig. 3 Structure of LSTM module

f_t , i_t , and O_t represent the outputs of the forget gate, input gate, and output gate at the current time step, respectively.

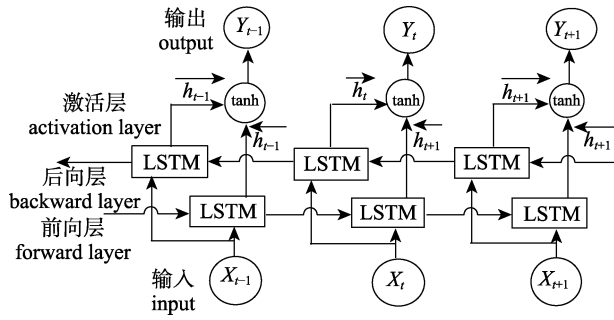


图 4 BiLSTM 模型结构

X_{t-1} 、 X_t 、 X_{t+1} 分别表示前一时刻、当前、后一时刻的输入数据; Y_{t-1} 、 Y_t 、 Y_{t+1} 分别表示前一时刻、当前、后一时刻的输出数据; h_{t-1} 、 h_t 、 h_{t+1} 分别表示前一时刻、当前、后一时刻的隐藏状态; tanh 表示双曲正切激活函数。

Fig. 4 Structure of BiLSTM module

X_{t-1} , X_t , and X_{t+1} represent the input data at the previous, current, and next time steps, respectively; Y_{t-1} , Y_t , and Y_{t+1} represent the output data at the previous, current, and next time steps, respectively; and h_{t-1} , h_t , and h_{t+1} represent the hidden states at the previous, current, and next time steps, respectively; tanh means the hyperbolic tangent activation function.

传播和后向传播两个方向上都应用 LSTM 单元,能够更全面地捕捉序列数据中的信息。在时间步 t 时,输入 X_t 传递到前向层和后向层中的 LSTM 单元,分别生成前向和后向的隐藏状态。随后,这些隐藏状态通过激活函数(tanh)生成输出 Y_t 。

本研究时间跨度和空间覆盖较大,将时空环境要素作为特征数据以环境时空变化对金枪鱼资源的影响。环境数量具有季节性变化,因此将月份作为候选变量之一纳入模型,以反映环境变量的季节性特征, LSTM 神经网络的记忆特性能够自动捕捉月份的周期性变化及其对 CPUE 的影响。基于 Bi-LSTM 进行 CPUE 预测本质上是循环神经网络的回归问题。以表征黄鳍金枪鱼 CPUE 的时空环境数据中的历史特征数据和当前特征数据作为网络输入, CPUE 值作为网络输出。通过与测量值的对比进行网络训练,从而建立特征数据与 CPUE 之间的映射关系,从而实现 CPUE 的预测。对于 CPUE 值,其在时间 t 时的特征值 $X(t)$ 可以表示为:

$$X(t) = \{\text{Year}, \text{Mon}, \text{Lon}, \text{Lat}, \text{Chl}, S_0, T_m, \text{DO}_m\} \quad (8)$$

式中, $X(t)$ 表示 t 时刻对应的年份(Year)、月份(Mon)、经度(Lon)、纬度(Lat)、叶绿素浓度(Chl)和盐度(S_0),以及水层 m 处的温度(T_m)和溶解氧浓

度(DO_m),例如 T_0 表示海表面的温度。为实现预测,将连续时间步长的 $X(t-n+1), \dots, X(t-1)$ 和 $X(t)$ 作为网络输入, t 时刻的 CPUE 即 $Y(t)$ 作为输出,其中 n 对应于输入层的步长。

归一化后的各变量作为 BiLSTM 模型的输入, CPUE 值为响应变量。模型的训练集、验证集和测试集按 6 : 2 : 2 比例划分,设置 epoch 为 300, batch size 为 50。训练中采用余弦退火学习率调度器,使得模型能够在初期快速收敛,并在训练后期避免震荡和过度更新^[35-36],学习率从 0.001 动态调整至最小值 0.00001。为防止过拟合,在 LSTM 层后加入 dropout (参数为 0.5)^[37],并使用 L2 正则化。早停法用于监控验证集损失,当连续 15 次迭代无改善时停止训练并恢复最佳权重。优化器采用 Nadam 以加快收敛并处理非平稳目标。预测模型在 PyCharm 中实现。模型的损失函数以均方误差(mean square error, MSE)表示,加入 L2 正则化后的损失函数计算如下:

$$\text{loss} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \lambda \sum_{j=1}^m w_j^2 \quad (9)$$

式中, n 表示损失值计算中的观察样本数量,包括训练集和验证集; Y_i 指样本序号为 i 的 CPUE 的预测值; $\hat{Y}_i$ 指样本序号为 i 的 CPUE 的值; m 表示模型中的总权重数目; w 表示模型的权重; w_j 表示第 j 个权重参数; λ 表示正则化强度的超参数,控制 L2 正则化项的权重。

1.4.2 模型性能评估 本研究采用平均绝对误差(mean absolute error, MAE)、均方根误差(root mean square error, RMSE)、平均绝对百分比误差(mean absolute percentage error, MAPE)和决定系数(coefficient of determination, R^2)作为评价模型性能的指标。在模型性能评估中, MAE、RMSE 和 MAPE 值越小,表明模型的稳定性越高; R^2 值越大,表示预测值与实际值的拟合度越高。各指标的计算公式如下:

$$\text{MAE} = \frac{1}{n} \times \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \times \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (11)$$

$$\text{MAPE} = \frac{\sum_{i=1}^n \left| \frac{\hat{Y}_i - Y_i}{Y_i} \right|}{n} \times 100 \quad (12)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2} \quad (13)$$

式中, n 表示测试集中的样本数量; $\bar{Y}$ 表示实际 CPUE 的平均值; Y_i 表示第 i 个样本的实际 CPUE; $\hat{Y}_i$ 表示第 i 个样本的预测 CPUE。

1.4.3 解释变量的贡献程度 SHAP (SHapley Additive exPlanations) 基于 Shapley 值解释机器学习模型的预测, 广泛用于深度学习的可解释性研究^[38]。每个 SHAP 值表示一个特征对模型输出的贡献, 通过计算特征贡献实现可视化, 揭示各解释变量的重要性。Shapley 值越大, 特征在模型中的重要性越高。通过联盟概念计算 SHAP 值(式 14 和 15)。在比较模型性能后, 选择最佳模型的数据集进行分析, 以识别对黄鳍金枪鱼渔场预测较为重要的变量。

$$\phi_i(y) = \frac{1}{N} \sum_{n=1}^N \phi_i^n \quad (14)$$

$$\phi_i^n = \hat{g}(y_{+i}^n) - \hat{g}(y_{-i}^n) \quad (15)$$

式中, $\phi_i(y)$ 表示预测实例 y 的 Shapley 特征值; ϕ_i^n 表示平均边际贡献; $\hat{g}(y_{+i}^n)$ 表示特征 i 存在的情况下的模型预测值; $\hat{g}(y_{-i}^n)$ 表示特征 i 不存在的情况下的模型预测值。

2 结果与分析

2.1 不同计算法的 CPUE 分布

对两种数据来源计算出的 CPUE 计算三分位点以区分渔场密度, 结果见表 1。两种 CPUE 的分布见图 5。结果表明, 基于不同数据集计算的 CPUE 不仅在数值范围上存在显著差异, 在相同地点表征的资源丰度的密集程度也不同。总体看, 基于渔捞日志的渔场南北分散, 而基于 VMS 的渔场主要集中在南部。

2.2 特征选择结果

基于 LASSO 回归的不同数据集中各变量的

表 1 CPUE 的三分位点计算结果

Tab. 1 Calculation results of the third quartiles of CPUE

CPUE	方法 1 method 1	方法 2 method 2
低值区 ($\leq Q1$) low-value interval	$0 < \text{CPUE} \leq 4.3$	$0 < \text{CPUE} \leq 5.2$
中值区 ($Q1, Q2$) mid-value interval	$4.3 < \text{CPUE} \leq 10.6$	$5.2 < \text{CPUE} \leq 11.8$
高值区 ($> Q2$) high-value interval	$\text{CPUE} > 10.6$	$\text{CPUE} > 11.8$

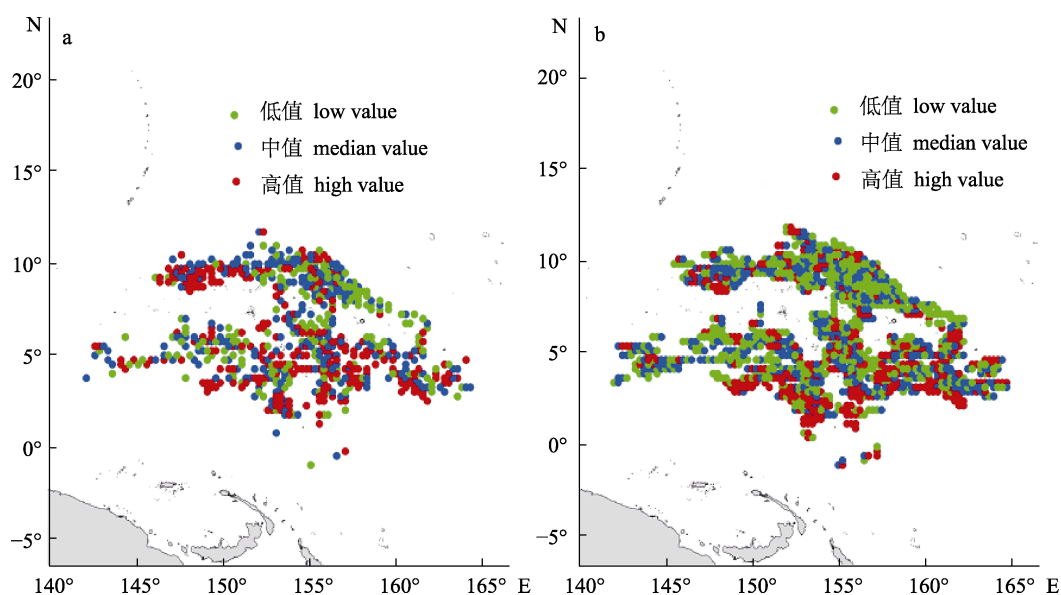


图 5 基于渔捞日志(a)和 VMS 数据集(b)的 CPUE 空间分布

Fig. 5 Distribution of estimated CPUE based on (a) logbook and (b) VMS datasets

系数见图 6。方法 1 数据集中 LASSO 回归系数为 0 的变量为: Year, Lat, DO₁₀₀, DO₂₀₀, DO₂₅₀, DO₃₀₀, T₀, T₂₀₀, T₂₅₀。方法 2 数据集中 LASSO 回归系数为 0 的变量为: Year, Lat, DO₀, DO₁₀₀, DO₂₅₀, DO₃₀₀, T₀, T₁₅₀, T₂₅₀。将 LASSO 回归系数为 0 的变量删除,剩下的变量的多重共线性诊断结果见表 2,各因子之间无多重共线性。通过特征选择的变量将用作 BiLSTM 建模。

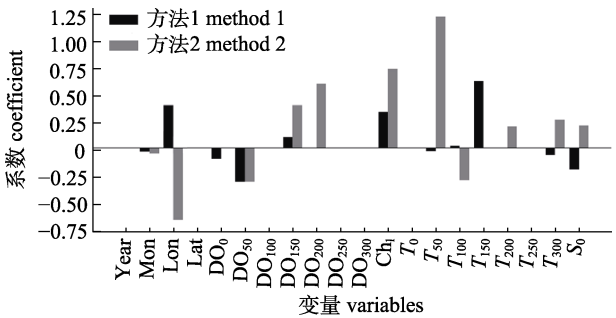


图 6 基于 LASSO 的变量筛选结果
Fig. 6 Variable screening result based on LASSO

2.3 BiLSTM 模型结果

模型训练完成后,训练集和验证集的损失函数变化如图 7 所示。基于不同计算方法的两种数据集的模型均经历了初始损失值略高,然后快速下降及稳定下降阶段。最终训练数据集和验证数据集的损失值相近,表明模型在训练集和验证集上都有良好的拟合效果,没有出现明显的过拟合现象。从损失函数看,方法 2 的验证集和训练集损失值几乎重合,其泛化能力可能比方法 1 更强,

表 2 变量的多重共线性诊断结果		
Tab. 2 Results of multicollinearity diagnosis		
变量 variable	方法 1 method 1	方法 2 method 2
Mon	1.862	1.597
Lon	1.499	1.382
S ₀	3.011	3.153
Chl	3.588	2.992
T ₅₀	3.038	1.874
T ₁₀₀	2.760	2.475
T ₂₀₀	/	3.281
T ₃₀₀	1.368	1.434
DO ₀	2.202	/
DO ₅₀	1.954	1.874
DO ₁₅₀	2.008	1.990
DO ₂₀₀	/	2.007

模型的鲁棒性更好。

2.4 模型预测误差比较

基于 BiLSTM 模型的不同数据集,验证数据集中预测的 CPUE 和实际 CPUE 如图 8 所示。可以观察到,基于渔捞日志的数据集(Method 1)的 CPUE 在低值与中值范围内的预测值和实际值之间的差异较小,而高 CPUE 值的预测误差较大。相比之下,基于 VMS 的数据集(Method 2)的预测总体上表现更好,实际值和预测值之间的差异较小。模型的性能指标见表 3。基于 VMS 的数据集的 RMSE、MAE 和 MAPE 较低, R²较高,表明模型拟合度更好,预测渔场的准确性更高。

2.5 变量的 SHAP 分析结果

为了解变量对黄鳍金枪鱼渔场预测的重要程

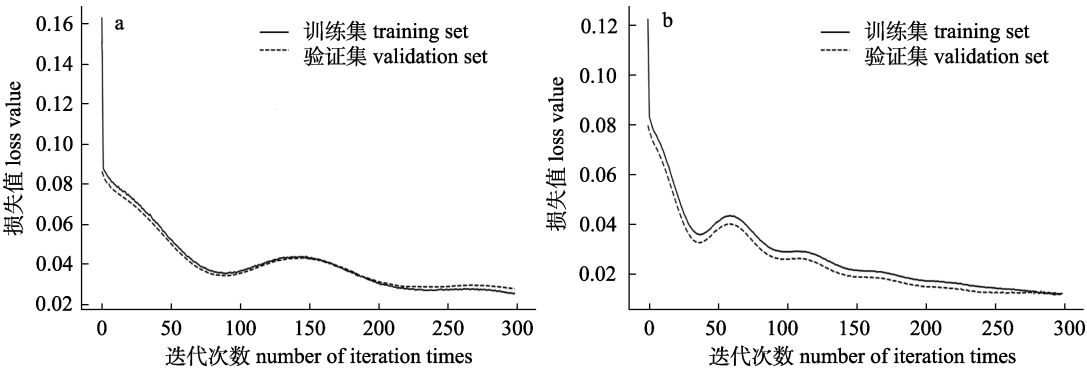


图 7 两种不同数据集在模型中的损失函数
a. 基于渔捞日志数据集的方法 1; b. 基于 VMS 数据集的方法 2.
Fig. 7 Loss functions of models based on two different datasets
a. Method 1 based on logbook dataset; b. Method 2 based on VMS dataset.

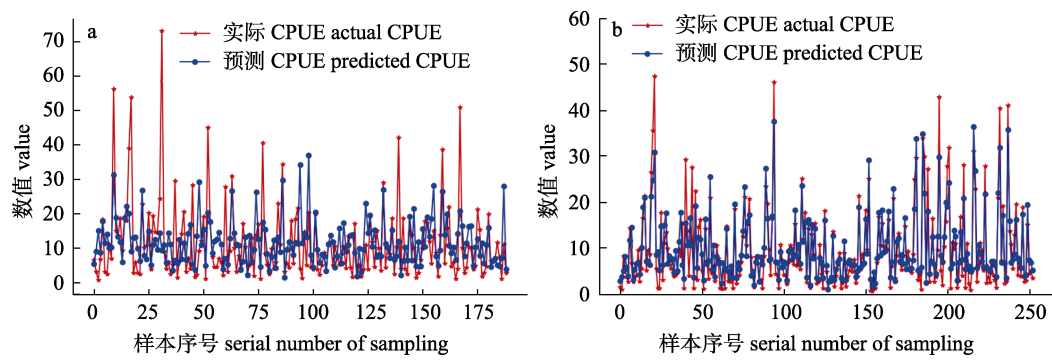


图 8 基于(a)渔捞日志和(b) VMS 数据集的 CPUE 预测值和实际值的比较
Fig. 8 Comparison of predicted and actual values of CPUE based on (a) logbook and (b) VMS datasets

表 3 模型的性能指标比较
Tab. 3 Comparison of model performance indicators

指标 indicator	方法 1 method 1	方法 2 method 2
均方根误差 RMSE	9.83	7.51
平均绝对误差 MAE	6.48	5.90
平均绝对百分比误差 MAPE	1.36	1.21
决定系数 R^2	0.26	0.42

度, 基于预测效果更好的数据集(即方法 2), 对 BiLSTM 渔场预测模型中的各变量进行 SHAP 分析, 结果如图 9 所示。结果显示, 在渔场预测模型中较为重要的变量依次是: Chl, DO_{200} , T_{300} , T_{200} , 以及 Lon 等。

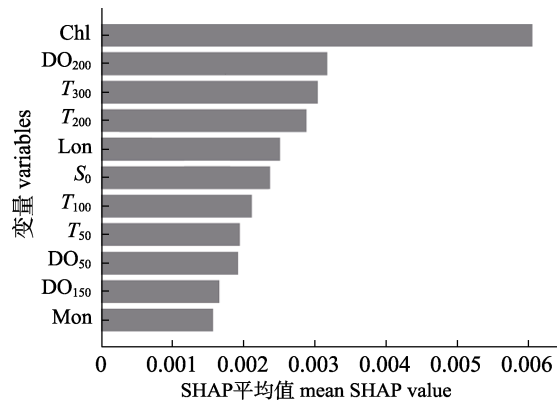


图 9 基于 VMS 数据集的各变量的 SHAP 值
Fig. 9 SHAP values of variables based on VMS dataset

2.6 基于 VMS 的 CPUE 的季度分布

基于 VMS 计算的 CPUE 各季度的空间分布如图 10 所示。黄鳍金枪鱼资源丰度有季节性的变化: 渔场在一二季度分布密集, 且明显多于三四季度。在第二季度时, 渔场密集区向东移动。第三季度 CPUE 低值区较多, 渔场密集区偏北, 在第

四季度偏南。

3 讨论

3.1 基于 VMS 计算 CPUE 的合理性

金枪鱼延绳钓渔业根据 VMS 估算 CPUE 更为合理。BiLSTM 模型结果表明, 在较小空间尺度内基于 VMS 估算的 CPUE 作为模型输入比基于渔捞日志的输入在预测渔场时有更高的精度。在利用渔业数据时应注意其空间分辨率, 渔船的空间分布可能会随着时间的推移而变化, 因此 CPUE 并非随机分布, 粗略计算可能出现解释中的偏差^[39]。延绳钓作业中, 作为主要捕捞对象的黄鳍金枪鱼在各天的捕捞量均大于 0, 但 VMS 推算的各网格中的捕捞量和努力量显示, 尽管某些网格中分配了钓钩, 但捕捞量或许为 0 (例如图 2 中的网格 E)。渔捞日志只简单记录单次作业的中心位置和渔获总量, 基于此计算 CPUE 时, 钓钩数量和渔获量被集中在每次作业的某一位置, 努力量和渔获量均未经过网格化处理, 而海洋遥感数据精确到网格中, 这些单元的时空环境特征影响了计算结果。在渔业的非独立调查中, VMS 记录的时间和空间分辨率相对于渔捞日志更高, 依据渔船在起绳期间的航程分配努力量到精细的网格中, 这符合延绳钓作业的实际情况, 延绳钓的努力量与渔获量往往是分布在投放作业的航迹线上^[40]。延绳钓捕捞的金枪鱼等海洋鱼类并非随机分布, 而是具有集群的特性, 渔获可能在某一较小的范围内被集中捕捞, 从而出现精细网格之间 CPUE 的较大差异。VMS 的计算反映了努力量与渔获量的实际分布情况, 根据渔船航程推算的 CPUE 在理论

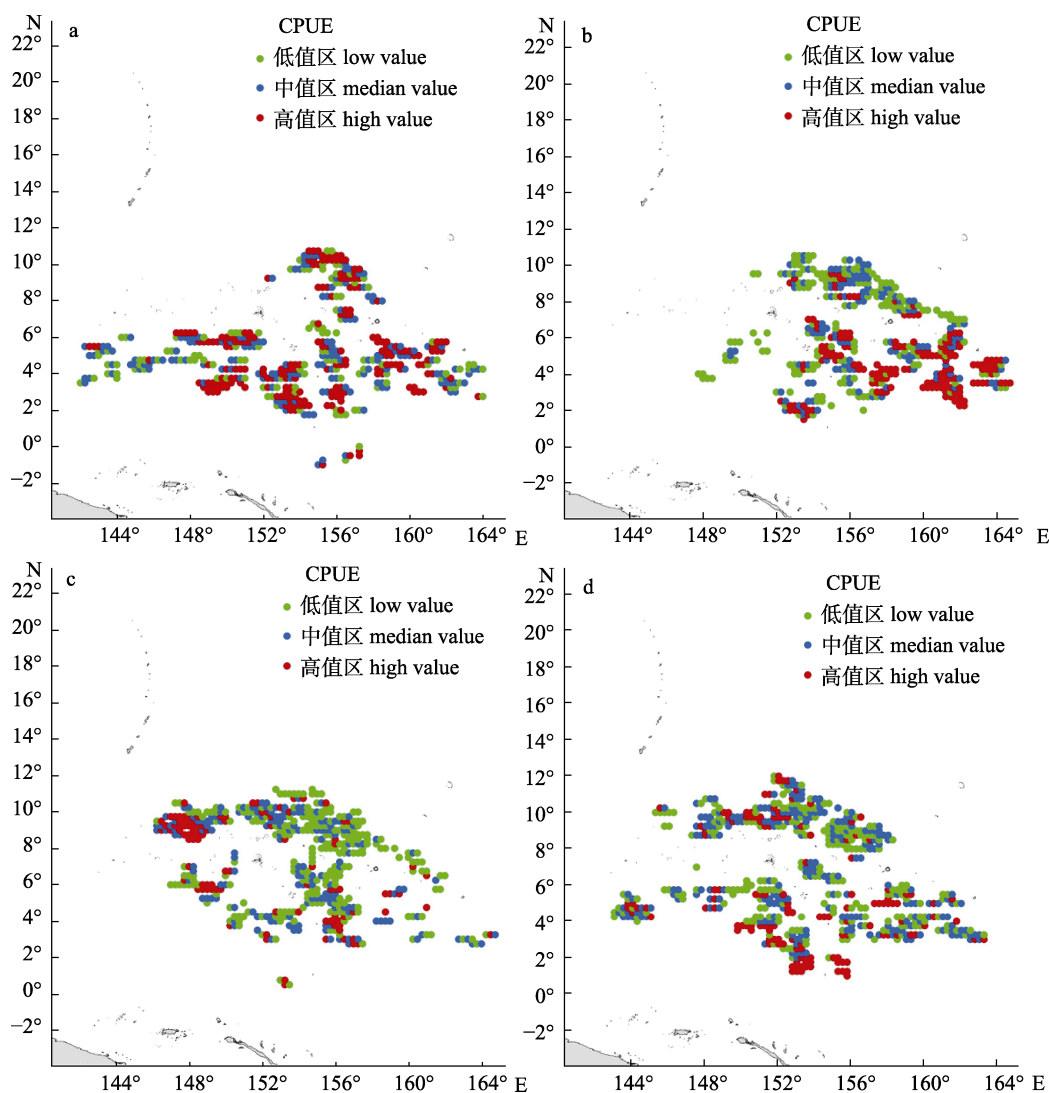


图 10 基于 VMS 的 CPUE 各季度空间分布

a. 春季; b. 夏季; c. 秋季; d. 冬季.

Fig. 10 Quarterly spatial distribution of CPUE based on VMS

a. Spring; b. Summer; c. Autumn; d. Winter.

上更加科学,因此基于此预测的渔场预报结果精度更高,相关的分析也更为可靠。本文的研究方法可以推广到其他渔业,在使用商业性渔业数据进行渔场预报等相关研究时,不仅需要考虑到不同模型对结果的影响,还要考虑作为模型输入的 CPUE 的统计方法不同对预测结果的影响,本文的比较分析方法可为选择最合适的 CPUE 计算方法提供参考。

CPUE 在本文中作为渔场资源丰度指标,并非绝对丰度的直接反映。作为资源丰度的近似反映,CPUE 在渔场预测的相关研究中通常被直接使用^[5-8],但仍存在一定局限性。CPUE 受到捕捞

效率、渔具参数、操作行为、时空环境等多种非生物学因素的影响,无法完全反映资源的真实分布和丰度。本文仅使用了渔业生产数据,未能结合独立的资源调查数据进行验证,未来的研究中应结合独立的资源调查数据(如声学或科学调查数据)以进一步验证模型在预测渔场的可靠性和适用性。

3.2 影响渔场预测的重要变量

本研究通过 LASSO 回归选取较为重要的变量以更好地建模,模型训练后通过分析各变量的 SHAP 平均值,得出各特征变量训练期间对渔场预测的贡献,为 BiLSTM 模型提供事后解释。研

究得出影响密克罗尼西亚海域黄鳍金枪鱼渔场预测的重要变量依次为: 叶绿素浓度, DO_{200} , T_{300} , T_{200} , Lon 等。叶绿素浓度反映了浮游生物量, 是海水中生产力水平的指标^[41], 影响鱼类的产量^[42]。叶绿素适宜的海域通常浮游生物充足, 小型鱼类丰富, 而黄鳍金枪鱼以小型鱼类为食, 因此叶绿素浓度可能直接影响其资源分布和种群丰度。海表面叶绿素浓度通常存在季节性波动^[43], 与之相关的生物资源也出现季节性变动的规律。金枪鱼耗氧量和标准代谢率超过大多数硬骨鱼^[44-45], 对

于溶解氧有较高的要求。溶解氧的限制可能会改变鱼类群落的组成, 从而影响金枪鱼的分布和捕捞量^[13,46]。溶解氧在各水层的垂直分布图显示, 溶解氧浓度随着水深增加呈现降低, 但在密克罗尼西亚海域南部, 溶解氧浓度仍保持着较高的值, 且明显高于北部海域(图 11)。温度影响黄鳍金枪鱼的产卵等活动^[47], 由于密克罗尼西亚海域地处热带, 浅水层的温度差值不大, 因此温度较大变化出现在更深的水层(图 12)。而研究海域内的南部, 即便深水层的温度也比北部更为适宜(图 12)。

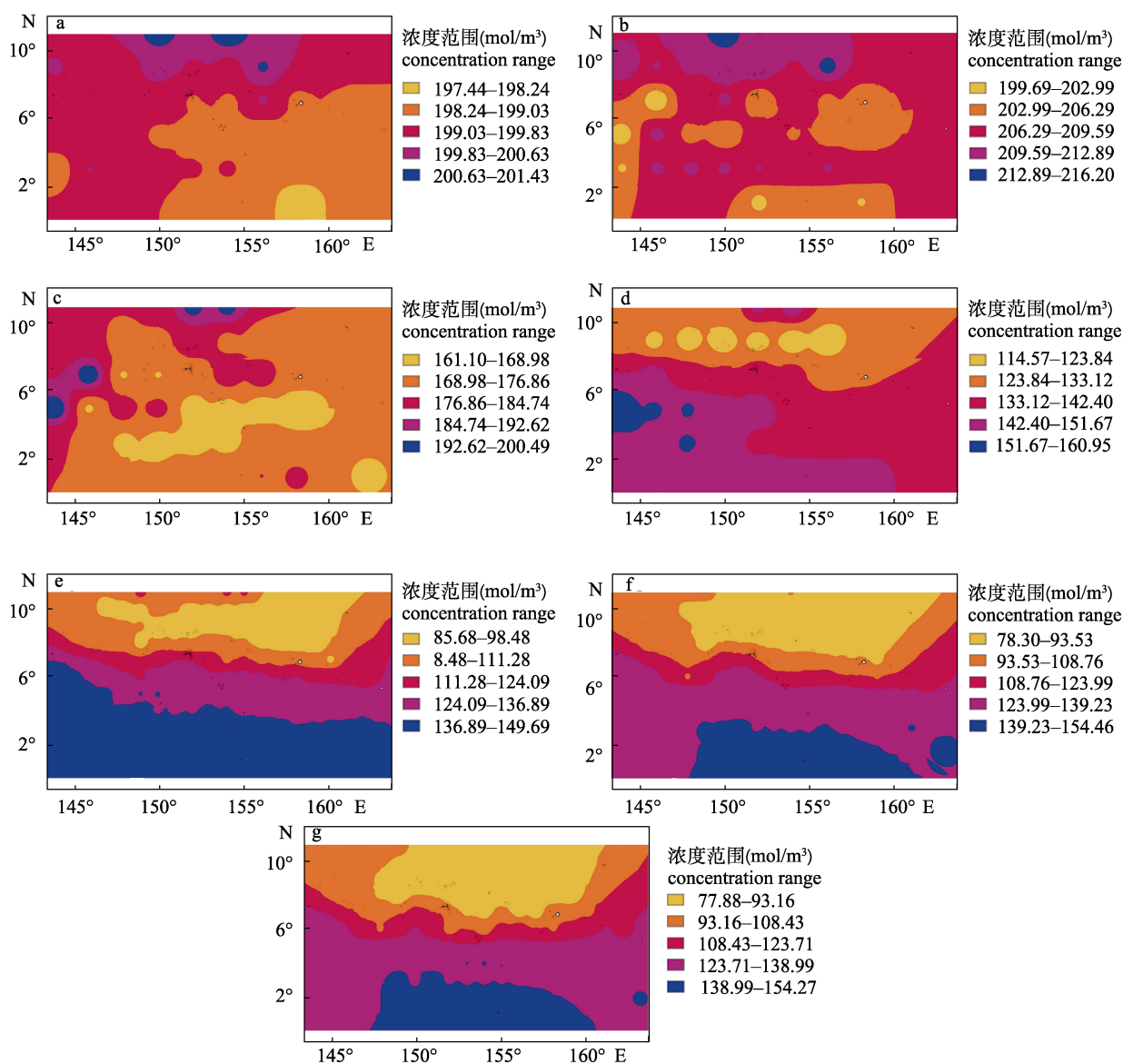


图 11 不同水层的溶解氧浓度分布

a. 0 m; b. 50 m; c. 100 m; d. 150 m; e. 200 m; f. 250 m; g. 300 m.

Fig. 11 Distributions of dissolved oxygen concentration at different water layers

a. 0 m; b. 50 m; c. 100 m; d. 150 m; e. 200 m; f. 250 m; g. 300 m.

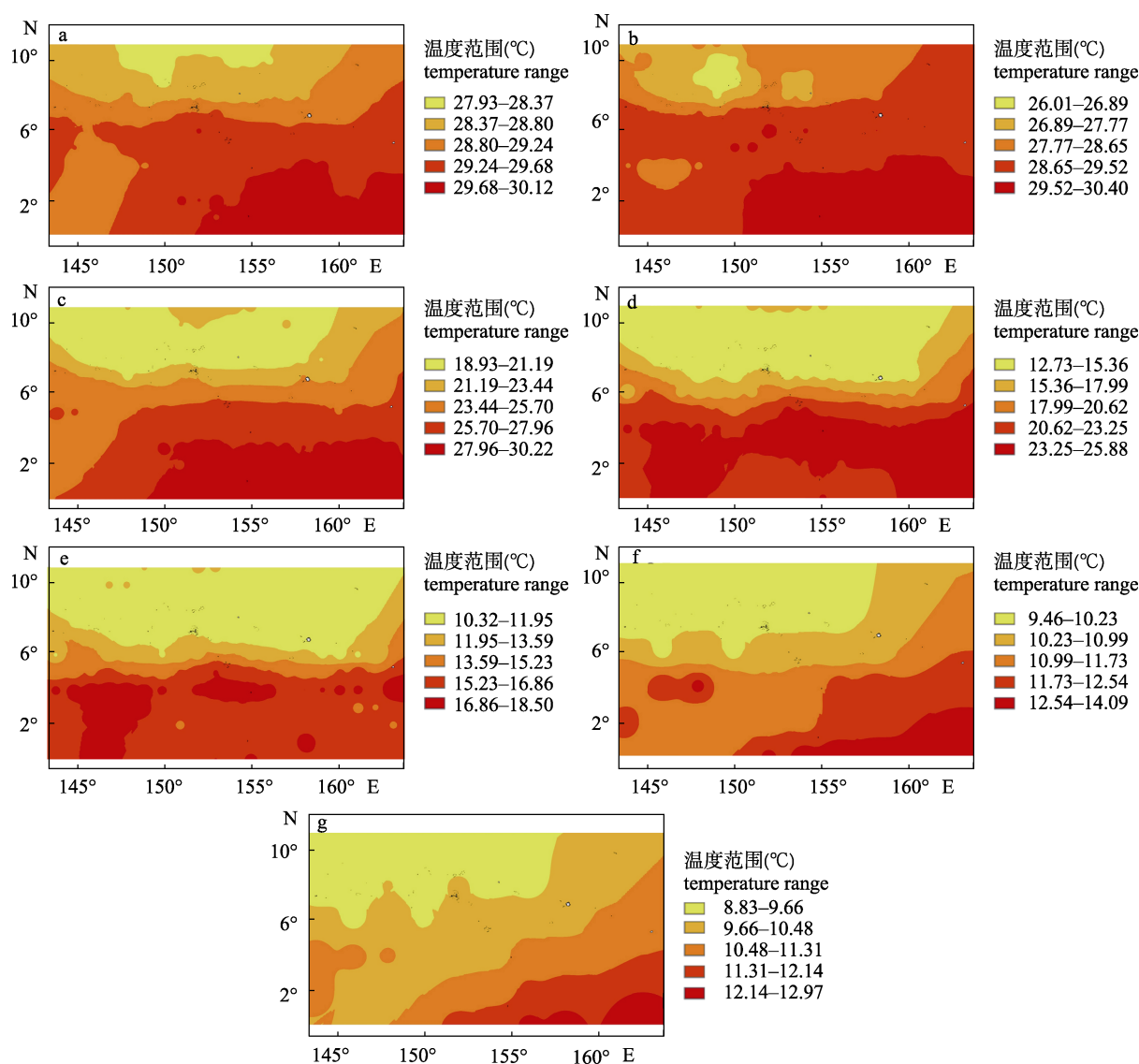


图 12 不同水层的温度分布

a. 0 m; b. 50 m; c. 100 m; d. 150 m; e. 200 m; f. 250 m; g. 300 m.

Fig. 12 Distributions of temperature at different water layers

a. 0 m; b. 50 m; c. 100 m; d. 150 m; e. 200 m; f. 250 m; g. 300 m.

水温和溶解氧浓度的垂直剖面图从理论上说明了黄鳍金枪鱼渔场偏南分布更为合理, 这也佐证基于 VMS 得出的渔场分布更为合理(图 5)。经度是影响黄鳍金枪鱼渔场的主要因素之一, 在其他研究也有类似结论^[48]。本文得出 200 m 和 300 m 为关键水层深度, 这或许与黄鳍金枪鱼的生态行为特征有关, 研究得出在中西太平洋深度 150 m 甚至 200 m 以深黄鳍金枪鱼渔获率较高^[49-50], 而杨胜龙等^[51]发现中西太平洋黄鳍金枪鱼的中心渔场集中在 17°C 等温线附近, 通常在 150~300 m 之间。基于本文的结论, 建议在密克罗尼西亚海域

捕捞黄鳍金枪鱼时将延绳钓钩沉降至 200 m 以深, 以提高渔获率。

3.3 Bi-LSTM 预测能力及误差分析

本研究得出基于 Bi-LSTM 的黄鳍金枪鱼 CPUE 预测效果较好。BiLSTM 通过结合前向和后向两个 LSTM 层, 可以同时捕捉时间序列中的双向依赖。这在渔情预测中非常重要, 因为 CPUE 等渔业数据往往受过去和未来条件的共同影响。双向依赖可以更准确地反映捕捞量随时间的变化趋势, 从而提升预测的准确性。渔情数据通常是时间序列数据, 涉及多年的捕捞记录。BiLSTM 能

够有效处理这些长时间序列,并捕捉到长时间跨度内的模式和依赖性,避免信息在长时间序列中丢失或减弱,这对于预测长期趋势或应对季节性变化尤其有效。双向信息的融合能够使 BiLSTM 模型在不同的渔业环境和条件下具备更好的泛化能力,有助于提高渔场预报的准确性。

延绳钓渔业中,CPUE 的大小与努力量有关,没有网格化统计努力量而直接将钓钩总数当作某一区域的努力量,导致 CPUE 值过小而计算错误。本研究在网格化统计努力量和渔获量后,得到 CPUE 值的波动范围在 0~100 尾/千钩之间。在实际数据波动较大的情况下,MAE 和 RMSE 显示出模型对大多数样本预测效果不错。在 CPUE 通常存在较高的随机性和复杂的非线性关系的渔业数据中,尤其在面对波动较大的数据时, R^2 低于 0.5 为合理情况^[52-53]。未来的研究可以通过加入更多影响金枪鱼 CPUE 的解释变量,或进一步优化模型,以进一步提升金枪鱼渔场预测模型的准确度。

4 结论

CPUE 常被用作相对资源丰度而在渔场预测模型的相关研究中作为响应变量。由于延绳钓渔船的数据来源有多种,需研究不同统计方法得出的 CPUE 作为预测模型的输入对渔场预报结果的影响。本研究的结果表明,利用影响黄鳍金枪鱼的一些关键时空环境要素,基于 BiLSTM 模型,可较为准确地预测金枪鱼资源丰度和渔场分布。本文的结果显示,不同数据来源的 CPUE 对渔场预测精度存在影响,以较小尺度研究地理跨度较大的延绳钓渔业时,利用 VMS 数据计算出的 CPUE 的数据集,在渔场预测模型中具有更高的精度。本文通过模型事后解释,揭示了影响延绳钓黄鳍金枪鱼渔场预测的重要因子和关键水层。本文初步研究了黄鳍金枪鱼 CPUE 的季节性时空分布规律以及因子的重要性,但环境变异性如何影响黄鳍金枪鱼资源变动的内在机理尚未进行探索。此外,黄鳍金枪鱼的分布还受摄食行为等制约^[54],未来可继续开展研究,可通过模拟鱼类对动态海洋景观中生物和非生物条件的空间异质性的反应来调查黄鳍金枪鱼群体的分布。CPUE 在

本文作为渔场资源丰度指标,并非绝对丰度的直接反映,由于本文仅使用了渔业生产数据,未能结合独立的资源调查数据进行验证,因此未来的研究中应结合独立的资源调查数据(如声学或科学调查数据)以进一步验证模型在预测渔场的可靠性和适用性。

参考文献:

- [1] Zhou C, Yang X, Yang B G, et al. Nutritional ecological niche of Pacific yellowfin tuna based on fatty acid analysis[J]. Journal of Fishery Sciences of China, 2023, 30(10): 1202-1213. [周成, 杨旭, 杨报国, 等. 基于脂肪酸分析的太平洋黄鳍金枪鱼营养生态位研究[J]. 中国水产科学, 2023, 30(10): 1202-1213.]
- [2] Kim J, Na H, Park Y G, et al. Potential predictability of skipjack tuna (*Katsuwonus pelamis*) catches in the Western Central Pacific[J]. Scientific Reports, 2020, 10: Article No. 3193.
- [3] Ram-Bidesi V. Regional arrangements for management and development of tuna in the Pacific Islands: The case study of Palau arrangement and the Federated States of Micronesia arrangement[C]//Proceedings of the Eleventh Biennial Conference of the International Institute of Fisheries Economics and Trade. Corvallis: International Institute of Fisheries Economics & Trade, 2002: 1-12.
- [4] Tanabe T. Feeding habits of skipjack tuna *Katsuwonus pelamis* and other tuna *Thunnus* spp. juveniles in the tropical western Pacific[J]. Fisheries Science, 2001, 67(4): 563-570.
- [5] 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.
- [6] Campbell R A. Constructing stock abundance indices from catch and effort data: Some nuts and bolts[J]. Fisheries Research, 2015, 161: 109-130.
- [7] Zhang H L, Ma Y C, Song H C, et al. Using structural equation modeling to investigate the influence of environmental factors on the CPUE of skipjack tuna (*Katsuwonus pelamis*) pair trawl fishery in Mauritania[J]. Journal of Fishery Sciences of China, 2024, 31(4): 465-475. [张鸿霖, 马有成, 宋厚成, 等. 基于结构方程模型研究环境因子对毛里塔尼亚双拖网 CPUE 的影响[J]. 中国水产科学, 2024, 31(4): 465-475.]
- [8] Zhang M, Song L M, Pan C, et al. Prediction on yellowfin tuna (*Thunnus albacares*) fishing ground in waters near the Marshall Islands based on SMOTETomek-RF[J]. Fisheries Oceanography, 2025, 34(2): e12704.
- [9] Tian S Q, Chen X J. Impacts of different calculating methods

- for nominal CPUE on CPUE standardization[J]. Journal of Shanghai Ocean University, 2010, 19(2): 240-245. [田思泉, 陈新军. 不同名义 CPUE 算法对 CPUE 标准化的影响[J]. 上海海洋大学学报, 2010, 19(2): 240-245.]
- [10] Lauridsen T L, Landkildehus F, Jeppesen E, et al. A comparison of methods for calculating catch per unit effort (CPUE) of gill net catches in lakes[J]. Fisheries Research, 2008, 93(1-2): 204-211.
- [11] Ebango Ngando N, Song L M, Cui H X, et al. Relationship between the spatiotemporal distribution of dominant small pelagic fishes and environmental factors in Mauritanian waters[J]. Journal of Ocean University of China, 2020, 19(2): 393-408.
- [12] Song L M, Li T L, Zhang T J, et al. Comparison of machine learning models within different spatial resolutions for predicting the bigeye tuna fishing grounds in tropical waters of the Atlantic Ocean[J]. Fisheries Oceanography, 2023, 32(6): 509-526.
- [13] Xu H, Song L M, Zhang T J, et al. Effects of different spatial resolutions on prediction accuracy of *Thunnus alalunga* fishing ground in waters near the cook islands based on long short-term memory (LSTM) neural network model[J]. Journal of Ocean University of China, 2023, 22(5): 1427-1438.
- [14] Emery T J, Noriega R, Williams A J, et al. Measuring congruence between electronic monitoring and logbook data in Australian Commonwealth longline and gillnet fisheries[J]. Ocean & Coastal Management, 2019, 168: 307-321.
- [15] Walsh W A, Ito R Y, Kawamoto K E, et al. Analysis of logbook accuracy for blue marlin (*Makaira nigricans*) in the Hawaii-based longline fishery with a generalized additive model and commercial sales data[J]. Fisheries Research, 2005, 75(1-3): 175-192.
- [16] Bigelow K, Musyl M K, Poisson F, et al. Pelagic longline gear depth and shoaling[J]. Fisheries Research, 2006, 77(2): 173-183.
- [17] Bastardie F, Nielsen J R, Ulrich C, et al. Detailed mapping of fishing effort and landings by coupling fishing logbooks with satellite-recorded vessel geo-location[J]. Fisheries Research, 2010, 106(1): 41-53.
- [18] Gaertner D, Dreyfus-Leon M. Analysis of non-linear relationships between catch per unit effort and abundance in a tuna purse-seine fishery simulated with artificial neural networks[J]. ICES Journal of Marine Science, 2004, 61(5): 812-820.
- [19] Yen K W, Lu H J, Chang Y, et al. Using remote-sensing data to detect habitat suitability for yellowfin tuna in the Western and Central Pacific Ocean[J]. International Journal of Remote Sensing, 2012, 33(23): 7507-7522.
- [20] Lopez J, Lennert-Cody C E, Maunder M N, et al. Developing alternative conservation measures for bigeye tuna in the eastern Pacific Ocean: A dynamic ocean management approach[C]//10th Meeting of the Scientific Advisory Committee. San Diego: IATTC, 2019: SAC-10-INF-D.
- [21] Song L M, Ren S Y, Zhang M, et al. Fishing ground forecasting models for yellowfin tuna (*Thunnus albacares*) in the tropical waters of the Atlantic Ocean based on ensemble learning[J]. Journal of Fishery Sciences of China, 2021, 28(8): 1069-1078. [宋利明, 任士雨, 张敏, 等. 基于集成学习的大西洋热带海域黄鳍金枪鱼渔情预报[J]. 中国水产科学, 2021, 28(8): 1069-1078.]
- [22] Nagano K, Yamamura O. Predicting catch of Giant Pacific octopus *Enteroctopus dofleini* in the Tsugaru Strait using a machine learning approach[J]. Fisheries Research, 2023, 261: 106622.
- [23] Xu H, Song L M, Li Y W, et al. Comparison of LSTM and QR models for predicting CPUE of albacore tuna in waters near the Cook Islands[J]. Fisheries Science, 2025, 91(2): 259-274.
- [24] Rakib A B, Rumky E A, Ashraf A J, et al. Mental healthcare chatbot using sequence-to-sequence learning and BiLSTM[C]//Proceedings of the 14th International Conference on Brain Informatics. Cham: Springer International Publishing, 2021: 378-387.
- [25] Ding P, Zou X R, Ding S Y, et al. Study on relationship between *Thunnus albacares* catches and climatic factors based on CNN-BiLSTM model[J]. South China Fisheries Science, 2024, 20(2): 19-26. [丁鹏, 邹晓荣, 丁淑仪, 等. 基于 CNN-BiLSTM 模型的黄鳍金枪鱼渔获量与气候因子关系研究[J]. 南方水产科学, 2024, 20(2): 19-26.]
- [26] Lutjeharms J R E, Allanson B R, Parker L. Frontal zones, chlorophyll and primary production patterns in the surface waters of the Southern Ocean south of Cape Town[J]. Elsevier Oceanography Series, 1986, 42: 105-117.
- [27] Yoshihara T. Distribution of fishes caught by the long line: II. Vertical distribution[J]. Nippon Suisan Gakkaishi, 1951, 16(8): 370-374.
- [28] Li G W, Gao X H, Xiao N W, et al. Estimation of soil organic matter content based on characteristic variable selection and regression methods[J]. Acta Optica Sinica, 2019, 39(9): 361-371. [李冠稳, 高小红, 肖能文, 等. 特征变量选择和回归方法相结合的土壤有机质含量估算[J]. 光学学报, 2019, 39(9): 361-371.]
- [29] Madbouly A I, Gody A M, Barakat T M. Relevant feature selection model using data mining for intrusion detection system[J]. International Journal of Engineering Trends and Technology, 2014, 9(10): 501-512.
- [30] Muthukrishnan R, Rohini R. LASSO: A feature selection

- technique in predictive modeling for machine learning[C]// Proceedings of the 2016 IEEE International Conference on Advances in Computer Applications (ICACA). Piscataway: IEEE, 2016: 18-20.
- [31] Abdulhafedh A. Comparison between common statistical modeling techniques used in research, including: Discriminant analysis vs logistic regression, ridge regression vs LASSO, and decision tree vs random forest[J]. Open Access Library Journal, 2022, 9(2): 1-19.
- [32] Akinwande M O, Dikko H G, Samson A. Variance inflation factor: As a condition for the inclusion of suppressor variable(s) in regression analysis[J]. Open Journal of Statistics, 2015, 5(7): 754-767.
- [33] Zhang Y F, Zhang S H, Liu Y X, et al. Radar HRRP sequence target recognition method of attention mechanism based stacked LSTM network[J]. Systems Engineering and Electronics, 2021, 43(10): 2775-2781. [张一凡, 张双辉, 刘永祥, 等. 基于注意力机制的堆叠 LSTM 网络雷达 HRRP 序列目标识别方法[J]. 系统工程与电子技术, 2021, 43(10): 2775-2781.]
- [34] Siامي-Namini S, Tavakoli N, Namin A S. The performance of LSTM and BiLSTM in forecasting time series[C]// Proceedings of the 2019 IEEE International Conference on Big Data (Big Data). Piscataway: IEEE, 2019: 3285-3292.
- [35] Mustafa E M, Elshafey M A, Fouad M M. Accuracy enhancement of a blind image steganalysis approach using dynamic learning rate-based CNN on GPUs[C]// Proceedings of the 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS). Piscataway: IEEE, 2019: 28-33.
- [36] Jepkoech J, Mugo D M, Kenduiwo B K, et al. The effect of adaptive learning rate on the accuracy of neural networks[J]. International Journal of Advanced Computer Science and Applications, 2021, 8(12): 736-751.
- [37] Ko B S, Kim H G, Oh K J, et al. Controlled dropout: A different approach to using dropout on deep neural network[C]// Proceedings of the 4th IEEE International Conference on Big Data and Smart Computing (BigComp). Piscataway: IEEE, 2017: 358-362.
- [38] Dou H, Zhang L M, Han F, et al. Survey on convolutional neural network interpretability[J]. Journal of Software, 2024, 35(1): 159-184. [窦慧, 张凌茗, 韩峰, 等. 卷积神经网络的可解释性研究综述[J]. 软件学报, 2024, 35(1): 159-184.]
- [39] Rose G A, Kulka D W. Hyperaggregation of fish and fisheries: How catch-per-unit-effort increased as the northern cod (*Gadus morhua*) declined[J]. Canadian Journal of Fisheries and Aquatic Sciences, 1999, 56(S1): 118-127.
- [40] Xu H, Song L M, Shen J R, et al. The relationship between the spatial-temporal distribution of albacore tuna CPUE and the marine environment variables in waters near the Cook Islands based on GAM[J]. Marine Science Bulletin, 2023, 42(4): 444-455. [许回, 宋利明, 沈介然, 等. 基于 GAM 的库克群岛海域长鳍金枪鱼 CPUE 时空分布与海洋环境的关系[J]. 海洋通报, 2023, 42(4): 444-455.]
- [41] Bell J D, Ganachaud A, Gehrke P C, et al. Mixed responses of tropical Pacific fisheries and aquaculture to climate change[J]. Nature Climate Change, 2013, 3(6): 591-599.
- [42] Gomez F, Montecinos A, Hormazabal S, et al. Impact of spring upwelling variability off southern - central Chile on common sardine (*Strangomera bentincki*) recruitment[J]. Fisheries Oceanography, 2012, 21(6): 405-414.
- [43] Xu H, Fujimori Y, Tomiyasu M, et al. Impact of different spatial scales on standardization of yellowfin tuna (*Thunnus albacares*) CPUE in waters near Micronesia[J]. Fisheries Science, 2025: 1-16.
- [44] Blank J M, Farwell C J, Morrisette J M, et al. Influence of swimming speed on metabolic rates of juvenile Pacific bluefin tuna and yellowfin tuna[J]. Physiological and Biochemical Zoology, 2007, 80(2): 167-177.
- [45] Fitzgibbon Q P, Seymour R S, Buchanan J, et al. Effects of hypoxia on oxygen consumption, swimming velocity and gut evacuation in southern bluefin tuna (*Thunnus maccoyii*)[J]. Environmental Biology of Fishes, 2010, 89(1): 59-69.
- [46] Cayré P, Marsac F. Modelling the yellowfin tuna (*Thunnus albacares*) vertical distribution using sonic tagging results and local environmental parameters[J]. Aquatic Living Resources, 1993, 6(1): 1-14.
- [47] Moore B R, Bell J D, Evans K, et al. Defining the stock structures of key commercial tunas in the Pacific Ocean I: Current knowledge and main uncertainties[J]. Fisheries Research, 2020, 230: 105525.
- [48] Wu Y X, He K M. Group normalization[C]// Proceedings of the 15th European Conference on Computer Vision (ECCV). Heidelberg: Springer-Verlag, 2018: 3-19.
- [49] Song L M, Shen Z B, Zhou J K, et al. Effects of environmental variables on catch rates of yellowfin tuna (*Thunnus albacares*) in waters near Cook Islands[J]. Journal of Shanghai Ocean University, 2016, 25(3): 454-464. [宋利明, 沈智宾, 周建坤, 等. 库克群岛海域海洋环境因子对黄鳍金枪鱼渔获率的影响[J]. 上海海洋大学学报, 2016, 25(3): 454-464.]
- [50] Schaefer K M, Fuller D W, Block B A. Movements, behavior, and habitat utilization of yellowfin tuna (*Thunnus albacares*) in the Pacific Ocean off Baja California, Mexico, determined from archival tag data analyses, including unscented Kalman filtering[J]. Fisheries Research, 2011, 112(1-2): 22-37.

- [51] Yang S L, Shi H M, Fan X M, et al. Spatial analysis of suitable habitat for yellowfin tuna in the tropical central and western Pacific Ocean[J]. *Journal of Agricultural Science and Technology*, 2022, 24(1): 183-191. [杨胜龙, 史慧敏, 范秀梅, 等. 热带中西太平洋黄鳍金枪鱼栖息水层空间分析[J]. *中国农业科技导报*, 2022, 24(1): 183-191.]
- [52] Ruta D, Cen L, Vu Q H. Deep bi-directional LSTM networks for device workload forecasting[C]// *Proceedings of the 15th Conference on Computer Science and Information Systems*. Piscataway: IEEE, 2020: 115-118.
- [53] Kugener G, Zhu Y C, Pangal D J, et al. Deep neural networks can accurately detect blood loss and hemorrhage control task success from video[J]. *Neurosurgery*, 2022, 90(6): 823-829.
- [54] Lan K W, Chang Y J, Wu Y L. Influence of oceanographic and climatic variability on the catch rate of yellowfin tuna (*Thunnus albacares*) cohorts in the Indian Ocean[J]. *Deep Sea Research Part II: Topical Studies in Oceanography*, 2020, 175: 104681.

Impact of different calculation methods for CPUE on the accuracy of fishing ground predictions in tuna longline fisheries

SONG Liming^{1,2}, XU Hui¹, QI Yukun¹, LU Canhuang³, SHEN Fengyuan⁴, DING Peng¹, LI Yuwei^{1,2*}, SHEN Jieran⁵

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. Zhejiang Ocean Family Co., Ltd., Zhoushan 316014, China;
4. East China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Shanghai 200090, China;
5. Liancheng Overseas Fishery Co., Ltd., Shenzhen 518035, China

Abstract: Accurately calculating catch per unit effort (CPUE) is a prerequisite for predicting fishery resource abundance and distribution. In tuna longline fisheries, the statistics of catch and effort are often influenced by spatial scale, and different sources of catch and effort data can result in biased CPUE estimates, which thereby potentially affect the forecast precision of fishing grounds. This study utilized fishery data obtained from longline vessels operating in waters near Micronesia, derived from both fishing logbooks and the VMS (vessel monitoring system), to calculate two types of CPUE for yellowfin tuna (*Thunnus albacares*). Using a BiLSTM (Bi-directional long-short-term memory) model, both types of CPUE were employed as inputs, alongside spatio-temporal factors and ocean remote sensing data, to develop a predictive model for yellowfin tuna fishing grounds. By SHAP (SHapley Additive exPlanations) analysis, key variables influencing the prediction of yellowfin tuna fishing grounds were identified. The spatio-temporal distribution characteristics of fishing grounds were determined based on the geographical distribution of different CPUE values. This study also evaluated the impact of different CPUE inputs on the accuracy of fishing ground prediction. The results showed that the BiLSTM model effectively predicts yellowfin tuna fishing grounds, with strong performance within the study area. Moreover, the model using CPUE calculated from VMS data demonstrated higher accuracy compared to the model built by using logbook data. Key variables influencing yellowfin tuna fishing ground predictions included chlorophyll-a concentration, dissolved oxygen concentration at 200 m, temperature at 300 m, temperature at 200 m, and longitude. Fishing grounds derived from fishing logbook were dispersed across the north and south, whereas those derived from VMS data were mainly concentrated in the southern region. Based on the vertical profile of temperature and dissolved oxygen, it was inferred that the fishing grounds identified by VMS data are more accurate. In the waters of Micronesia, yellowfin tuna fishing grounds were more densely distributed in the first and second quarters. This study demonstrates that VMS-based CPUE can be calculated for longline fisheries and effectively used for fishing ground prediction. Based on this study, it is recommended to record catch locations individually using the VMS method, when investigating tuna fishing grounds in longline fisheries at smaller spatial scales. This allows for grid-based CPUE calculations, thereby improving the accuracy of fishing ground prediction.

Key words: longline fishing; *Thunnus albacares*; BiLSTM; CPUE; fishing ground prediction

Corresponding author: LI Yuwei. E-mail: ywli@shou.edu.cn