

文章编号: 1674-5566(2025)01-0153-11

DOI: 10.12024/jsou.20241104680

基于LSTM模型的中西太平洋鲷栖息地预测

周成^{1,2,3,4}, 周想¹, 胡媛媛¹, 刘力文¹

(1. 上海海洋大学 海洋生物资源与管理学院, 上海 201306; 2. 国家远洋渔业工程技术研究中心, 上海 201306; 3. 大洋渔业资源可持续开发教育部重点实验室, 上海 201306; 4. 农业农村部大洋渔业可持续利用重点实验室, 上海 201306)

摘要: 为了解决传统栖息地预测模型中无法捕捉具有时间序列信息的环境因子对金枪鱼空间分布滞后影响的不足。采用2021—2024年金枪鱼围网渔捞日志数据, 通过构建滞后天数为1、5、10、15 d的长短期记忆(Long-short term memory, LSTM)神经网络模型, 分别对单位捕捞努力量渔获量(Catch per unit effort, CPUE)和经纬度进行了预测。研究表明, 滞后10 d的模型精度最高, 其均方误差(Mean square error, MSE)为0.018 7, 平均绝对误差(Mean absolute error, MAE)为0.077 6, 表明鲷空间分布受过去短期内环境累积效应的影响。通过对最佳模型进行验证, 结果表明预测纬度与实际纬度之间的 R^2 为0.97, 预测经度与实际经度之间的 R^2 为0.65, 说明空间分布预测范围与实际基本吻合。为揭示鲷栖息地特征及其生态过程的动态机制提供了新的理解, 同时为中西太平洋鲷围网渔业的科学管理提供了重要参考依据。

关键词: 鲷; 栖息地预测; 长短期记忆模型; 中西太平洋

中图分类号: S 931.1

文献标志码: A

鲷(*Katsuwonus pelamis*)广泛分布于太平洋、大西洋和印度洋^[1], 其捕捞量常年位居全球单种海洋捕捞量的前列^[2-3], 仅次于日本鰹(*Engraulis japonicus*)和狭鳕(*Theragra chalcogramma*), 被认为是最具商业价值的金枪鱼种类之一^[4]。鲷的主要捕捞方式为金枪鱼围网, 其产量约占鲷总捕捞量的60%以上^[5]。理解鲷栖息地的动态分布有利于识别渔场热点并优化围网作业布局。传统的栖息地分布研究通常基于静态模型, 通过匹配即时环境因子与渔业数据来预测鱼类分布。例如, ZAINUDDIN等^[6]基于广义加性模型(Generalized additive model, GAM)和广义线性模型(Generalized linear model, GLM)预测了西北太平洋长鳍金枪鱼(*Thunnus alalunga*)的渔场分布; 许回等^[7]和徐国强等^[8]通过构建GAM模型, 揭示了海洋环境因子对金枪鱼栖息地分布的影响; 周为峰等^[9]利用贝叶斯分类器在南海黄鳍金枪鱼(*Thunnus albacares*)的栖息地预测中获得

了较为准确的结果; 张聪等^[10]利用集成模型对中西太平洋黄鳍金枪鱼渔场进行预测, 克服了环境因子和模型单一的局限并取得良好效果。

由于鲷的分布对环境变动的响应存在一定的时滞效应^[11-13], 通过动态环境变化的过程研究鲷栖息地的动态分布, 能够更精准地捕捉环境因子与生物响应之间的关系。近年来, 部分研究开始关注金枪鱼受环境因子影响的滞后效应, 例如, 张健等^[14]探究了黄鳍金枪鱼和鲷渔获量与海洋尼诺指数(Oceanic Niño index, ONI)之间的滞后关系; 杜艳玲等^[15]基于ConvLSTM-CNN模型对太平洋长鳍金枪鱼渔场进行了预测, 结果显示在考虑时滞效应后, 渔场预测的精度显著提升。这些方法克服了传统模型未能有效结合时间序列环境数据的局限性, 但目前对于鲷空间分布受环境因子滞后影响的研究鲜有报道。

为此, 本研究基于2021—2024年金枪鱼围网渔捞日志数据, 结合多个水层的时间序列环境因

收稿日期: 2024-11-01

修回日期: 2024-12-15

基金项目: 国家重点研发计划(2023YFD2401301); 农业农村部全球渔业资源调查监测评估(公海渔业资源综合科学调查)专项(D-8025-23-1003)

作者简介: 周成(1987—), 男, 副教授, 博士, 研究方向为金枪鱼渔业。E-mail: c-zhou@shou.edu.cn

通信作者: 胡媛媛, E-mail: hyywork97@163.com

版权所有 ©《上海海洋大学学报》编辑部(CC BY-NC-ND 4.0)

Copyright © Editorial Office of Journal of Shanghai Ocean University (CC BY-NC-ND 4.0)

<http://www.shhydx.com>

子^[16-18],构建不同滞后天数的长短期记忆(Long-short term memory, LSTM)神经网络模型,对单位捕捞努力量渔获量(Catch per unit effort, CPUE)及经纬度进行了预测。研究结果能够为解析鳀栖息地特征及其生态过程的动态机制提供新的理解,并为中西太平洋鳀围网渔业的科学管理提供重要的参考依据。

1 材料与方法

1.1 数据来源

渔业生产数据来源于国家远洋渔业数据中心提供的2021年11月—2024年1月中国金枪鱼围网渔船捕捞日志。日志记录了包括作业时间(年、月、日)、作业位置(经纬度)、捕捞网次、捕捞类型和渔获产量等信息,作业范围覆盖15°S~15°N, 130°E~145°W。

太平洋鳀通常栖息于约100 m的水层,其种群丰度在0~45 m水层达到最大值^[19-20]。综合考虑鳀日间垂直移动的栖息深度范围^[21-22],选取了0、50、100、150和200 m共5个水层的环境数据进行分析。生物环境数据包括初级生产力(Primary productivity, PP)和叶绿素a质量浓度(Chlorophyll a, *Chl.a*),非生物环境数据包括温度(Sea temperature, ST)、盐度(Salinity, SS)、溶解氧浓度(Dissolved oxygen, DO)、海表面高度(Sea surface height, SSH)、混合层厚度(Mixed layer depth, MLD)、东向海流速度(Eastward current velocity, ECV)和北向海流速度(Northward current velocity, NCV)。环境数据来源于欧盟哥白尼海洋环境监测服务中心(Copernicus marine environment monitoring service, 网址: <https://data.marine.copernicus.eu/products>),时间范围为2021年11月1日至2024年1月31日,空间分辨率为0.25°×0.25°,时间分辨率为1 d。其中,地转海流速度^[23](Geostrophic current velocity, GCV)由东向和北向海流速度计算得出,计算公式为

$$G_i = \sqrt{E_i^2 + N_i^2} \quad (i = 0, 50, 100, 150, 200) \quad (1)$$

式中: G_i 为 i 水层的 GCV; E_i 为 i 水层的 ECV; N_i 为 i 水层的 NCV; i 为不同水层。

1.2 数据预处理

由于捕捞自由鱼群过程中因捕捞技术的限制导致鱼群逃逸而产生空网,本研究将渔捞数据中的零产值记录剔除。为进一步避免异常产量

数据对模型结果的影响,对剩余数据中产量为正的数据进行清洗,根据箱线图识别异常值并剔除。

本研究数据的时间分辨率为1 d,空间分辨率为0.25°×0.25°,研究区域被划分为41 261个空间单元。基于相同日期和作业经纬度,通过最小欧氏距离原则匹配环境数据和渔捞数据,将作业站点对应到相应的渔区单元,并将站点坐标转换为渔区单元的中心点坐标。将日期和经纬度相同的样本行合并,计算每个渔区的单日总产量和作业总网次,同时取匹配环境数据的均值。

构建包含时间序列样本量 n 、时间步长 d 和解释变量 N 等3个维度的样本数组。在0.25°×0.25°的空间尺度下,设定时间步长为 d 天,即当前日期为 t ,分别匹配日期为 $t-d$ 、 $t-d+1$ 、 \dots 、 $t-1$ 的环境数据,获得维度为 $n \times d$ 的矩阵。将这些矩阵按解释变量拼接,生成符合LSTM模型输入参考格式的样本数组,其维度为 $n \times d \times N$ 。

1.3 CPUE 计算

使用基于网次的CPUE作为鳀种群相对丰度的指标,其计算方法:

$$Y_n = \frac{C_n}{E_n} \quad (n = 1, 2, 3, 4, \dots) \quad (2)$$

式中: Y_n 为第 n 行捕捞记录中的CPUE, t /网; C_n 为第 n 行捕捞记录中的鳀产量, t ; E_n 为第 n 行的捕捞网次,网。

1.4 变量筛选

采用海洋环境变量分析鳀的丰度及其时空分布变化,针对解释变量之间可能存在的相互依赖关系,考虑到多重共线性可能引起的参数估计不确定性,以及由此导致的估计值方差和协方差增加,从而影响模型预测效果,在数据预处理阶段对解释变量进行了多重共线性检验和特征筛选,以避免多重共线性对模型准确性的不利影响。

岭回归(Ridge regression, RR)通过在损失函数中加入 λ_2 正则项来限制模型系数过大^[24],套索回归(Least absolute shrinkage and selection operator, LASSO)在传统线性回归的基础上添加 λ_1 正则项,通过压缩系数来剔除系数为零的变量,从而选择出有效的特征变量^[25]。弹性网络(Elastic net, EN)模型综合RR和LASSO的优

点^[26],通过交叉验证选择最优 λ 值,有效解决多重共线性问题并筛选关键变量。

构建 X 矩阵和 Y 矩阵作为EN模型的输入,其中 X 矩阵包含所有自变量, Y 矩阵涵盖CPUE的时间序列和空间坐标。为消除不同量纲对模型的影响,对 X 和 Y 矩阵进行归一化处理。采用归一化处理后的数据构建EN模型,通过交叉验证确定最佳的 λ 参数,保留最优参数下系数不为零的特征变量作为构建LSTM模型的解释变量。

1.5 LSTM 模型

1.5.1 CPUE 时空分布与海洋环境特征的 LSTM 模型构建

LSTM 在递归神经网络 (Recurrent neural network, RNN) 的基础上引入了遗忘门、输入门和输出门 3 种“门”机制,用于判断、筛选和更新输入信息,从而保留长距离依赖信息并实现信息流的平稳传输^[27-29]。本研究基于时间序列将数据集划分为训练集、验证集和测试集以构建 LSTM 模型。训练集用于拟合数据并优化模型权重,通过梯度下降算法最小化误差,占总数据量的 80%,时间范围为 2021 年 11 月至 2024 年 1 月;验证集为训练集的 20%,用于评估模型性能和调优超参数;测试集为剩余的 20%,用于评估模型的预测能力,并将预测值与实际值进行对比分析。

为评估不同时间步长的 LSTM 模型在 $0.25^\circ \times 0.25^\circ$ 空间尺度下的预测性能,本研究构建了 4 种时间步长的 LSTM 模型。初级生产力是金枪鱼渔场形成的重要驱动因素之一,受多种海洋环境因子的综合影响^[30-31]。初级生产力较高的区域能促进浮游植物的繁殖,从而吸引以其为食的生物,并进一步吸引金枪鱼群聚索饵。从初级生产力的富集到高生产力的形成通常存在 2~3 周的滞后时间^[32],因此,模型分别设定 1、5、10 和 15 d 等 4 种滞后时间步长。自变量由 EN 模型筛选获得,因变量为渔区单元中心点的经纬度和 CPUE。LSTM 模型结构包含 4 层:第一层为 LSTM 层,用于特征提取;第二层和第三层为隐藏层,用于特征分类和维度转换;最后一层为输出层,用于生成预测结果。LSTM 层和隐藏层采用 ReLU 激活函数,以增强对非线性特征的处理能力并避免梯度消失问题;输出层则使用线性回归函数输出预测的经度、纬度和 CPUE。优化算法选择适应性梯度算法 (Adaptive gradient algorithm,

Adadelta),通过对梯度平方的指数衰减移动平均动态调整学习率,从而提高训练效率和模型性能。

采用网格搜索与调优方法优化超参数设置。训练过程中,迭代次数设定为 100,每批样本大小为 32,并结合早停机制与学习率调度器提高训练效率。当连续 10 个训练周期 (patience=10) 验证集性能无改善时,提前停止训练;若验证损失在 5 个周期 (patience=5) 内未改善,则将学习率减小为当前值的 20%。针对模型的前三层神经元数量、学习率、正则化范数 (L2) 以及随机失活率进行了参数调优,具体参数如下,神经元数量:16、32、64;学习率:0.001、0.000 1;正则化范数:0.001、0.000 1;随机失活率:0.1、0.2。通过随机网格搜索抽取的参数组合,共计 432 种超参数组合,最终筛选出最优参数设置。在模型评估中,选择使均方误差 (Mean square error, MSE) 和平均绝对误差 (Mean absolute error, MAE) 最小的时间步长作为最终模型,并在测试集上进行预测验证。同时,通过渔场重叠概率分析评估 LSTM 模型的预测性能,以进一步验证模型的效率。

1.5.2 LSTM 模型的性能评估

为评估 LSTM 模型的性能,选择 MSE 作为损失函数。MSE 能够量化模型预测值与真实值之间的差距,其平滑的函数曲线有助于实现模型的稳定收敛。MSE 值越小,表明模型的预测精度越高^[33-34]。此外,采用 MAE 作为性能评价指标^[35-36],MAE 通过计算预测值与实际值之间差异的绝对值平均值,更直观地表征数据的正常分布情况。MAE 对异常值的敏感性较低,当 MAE 值越小时,表明模型整体表现越优^[37]。两种指标的计算公式:

$$E_{MS} = \frac{1}{m} \sum_{i=1}^m (Y_i - \hat{Y}_i)^2 \quad (3)$$

$$E_{MA} = \frac{1}{m} \sum_{i=1}^m |(Y_i - \hat{Y}_i)| \quad (4)$$

式中: E_{MS} 为 Y 矩阵变量真实值和预测值的均方误差 MSE; E_{MA} 为 Y 矩阵变量和预测值的平均绝对误差 MAE; m 为训练样本数量; Y_i 为 Y 矩阵变量真实值; \hat{Y}_i 为 Y 矩阵变量预测值。

通过对 4 种不同时间步长的 LSTM 模型进行训练迭代和超参数优化,并综合 MSE 和 MAE 对模型性能进行评估,将 MSE 作为损失函数,将

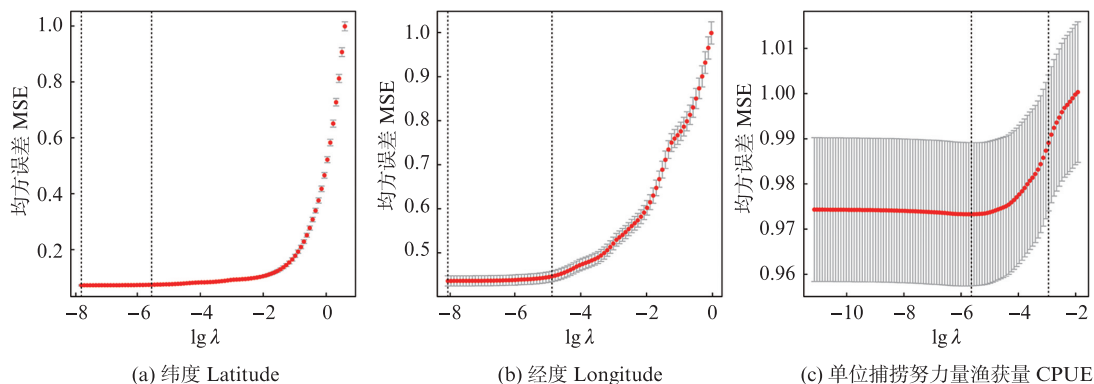
MAE 作为评价指标,选取 2 项指标均达到最小值的迟滞时间作为最优模型。最终,利用最优模型对 2022 年鳀 CPUE 值及其经纬度分布进行预测。

2 结果

2.1 筛选环境特征

采用 LSTM 模型进行时滞分析,在自变量中不再考虑年份、月份和日期。EN 模型的输入矩阵 X 由 32 个自变量组成,涵盖 $Chl.a-0$ 至 $Chl.a-200$ 、 $PP-0$ 至 $PP-200$ 、 $ST-0$ 至 $ST-200$ 、 $SS-0$ 至 $SS-200$ 、 $DO-0$ 至 $DO-200$ 、 $GCV-0$ 至 $GCV-200$ 、MLD 及 SSH,而 Y 矩阵则包含 Lat、Lon 和 CPUE 等 3 个因变量。

设定 α 值为 0.5,分别以 Lat、Lon 和 CPUE 作为因变量,构建基于海洋环境特征的 EN 模型。在变量筛选过程中,重要性较低的变量会优先被压缩至 0 并从模型中剔除。通过交叉验证确定最优 λ 值,选择使平均交叉验证误差最小的 λ 值作为最终解。图 1 显示了不同因变量的交叉验证误差曲线及其标准差曲线。根据 3 个因变量分别筛选出来的重要环境特征,共同重要的环境特征有 26 个,分别为: $DO-0$ 、 $DO-100$ 、 $DO-200$ 、 $Chl.a-50$ 、 $Chl.a-100$ 、 $Chl.a-150$ 、 $Chl.a-200$ 、 $PP-0$ 、 $PP-50$ 、 $PP-100$ 、 $PP-200$ 、 $SS-50$ 、 $SS-100$ 、 $SS-150$ 、 $SS-200$ 、 $ST-50$ 、 $ST-100$ 、 $ST-150$ 、 $ST-200$ 、 $GCV-0$ 、 $GCV-50$ 、 $GCV-100$ 、 $GCV-150$ 、 $GCV-200$ 、MLD 和 SSH。



λ 为模型的正则化参数,图中两条垂直的虚线代表模型给出的两个 λ 值。左边第一条垂直线代表达到最小平均交叉验证误差的 λ 值,第二条垂直线代表使平均交叉验证误差在最小平均交叉验证误差的标准差范围内的最大 λ 值。

λ is the regularization parameter of the model. The two vertical dashed lines in the plot represent the two λ values provided by the model. The first vertical line on the left represents the λ value that minimizes the average cross-validation error, while the second vertical line represents the maximum λ value for which the average cross-validation error falls within the standard deviation range of the minimum average cross-validation error.

图 1 不同因变量的交叉验证曲线和标准差曲线

Fig. 1 Cross-validation curves and standard deviation curves for different dependent variables

2.2 不同时间步长的最优 LSTM 模型

EN 模型共筛选出的 26 个海洋环境特征,分别构建滞后 1、5、10 和 15 d 的 CPUE 时空分布与海洋环境特征的 LSTM 模型。每个模型经过 30 次循环训练,每次循环包含 100 次迭代,由于早停机制的存在,通常不需迭代至 100 次就可结束。通过 MSE 计算模型损失,并使用 MAE 评估模型性能。根据循环训练过程中 MSE 和 MAE 的最低值确定各时间步长下 LSTM 模型的最优超参数组合,训练结果如表 1 所示。结果表明在 4 种时间步长的模型中,神经元数量和正则化范数范围较小且学习率固定为 0.001 时,模型效果更佳。通过比较不同时间步长下的最优模型,并根据 MSE 和 MAE 值进行排序,结果显示,基于滞后

10 d 的海洋环境数据构建的 LSTM 模型具有最高的预测精度和最佳的性能。

2.3 最优 LSTM 模型拟合情况

图 2 为 4 种时间步长的最优 LSTM 模型的 MSE 和 MAE 损失曲线。总体来看,训练集和测试集的损失值均呈下降趋势,并在迭代至 50 次后逐渐趋于稳定,损失值的下降幅度减小,最终收敛至固定数值。随着迭代次数的增加,训练集和测试集的误差显著减小。从损失曲线可以看出,4 种时间步长的最优模型均未出现欠拟合或过拟合现象。相比其他时间步长的 LSTM 模型,滞后 10 d 的模型在训练集和验证集中的误差较小,收敛速度较快,并在约 50 次迭代后逐渐稳定,表明该模型性能优越。

表1 不同时间步长LSTM的最优超参数、损失和MAE
Tab. 1 Optimal hyperparameters, Loss and MAE of LSTM with different time steps

时间步长 Time step/d	滞后时间区间 Lag time interval/d	神经元数量 Number of neurons	学习率 Learning Rate	正则化范数 L2	随机失活概率 Dropout probability	损失 Loss	平均绝对误差 MAE
1	0~1	64, 64, 16	0.001	0.000 1	0.1, 0.2	0.022 9	0.097 9
5	0~5	64, 32, 64	0.001	0.000 1	0.1, 0.2	0.019 5	0.082 3
10	0~10	32, 32, 64	0.001	0.000 1	0.1, 0.1	0.018 7	0.077 6
15	0~15	64, 16, 64	0.001	0.000 1	0.1, 0.2	0.018 9	0.079 0

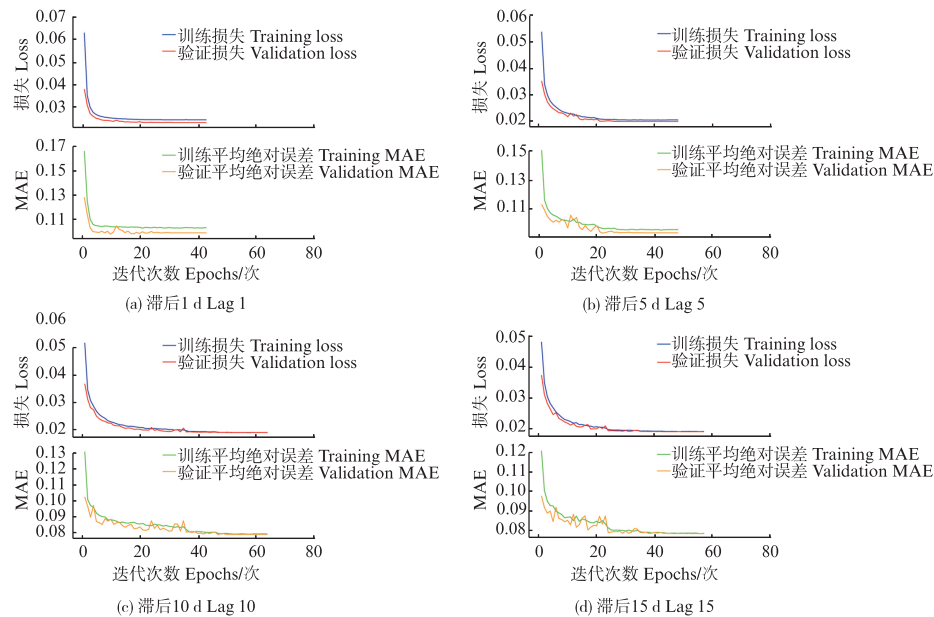


图2 不同时间步长LSTM的最佳模型损失函数曲线图

Fig. 2 Loss function curve of the optimal model LSTM with different time steps

2.4 LSTM模型预测精度

本研究基于滞后10 d海洋环境数据构建了用于鲷栖息地预测的LSTM模型,并对2022年鲷栖息地的时空分布进行了预测。经过反归一化处理,将预测结果与实际CPUE和空间坐标进行比较。图3显示了基于滞后10 d时间步长的LSTM模型预测结果与实际值的时间序列对比图。

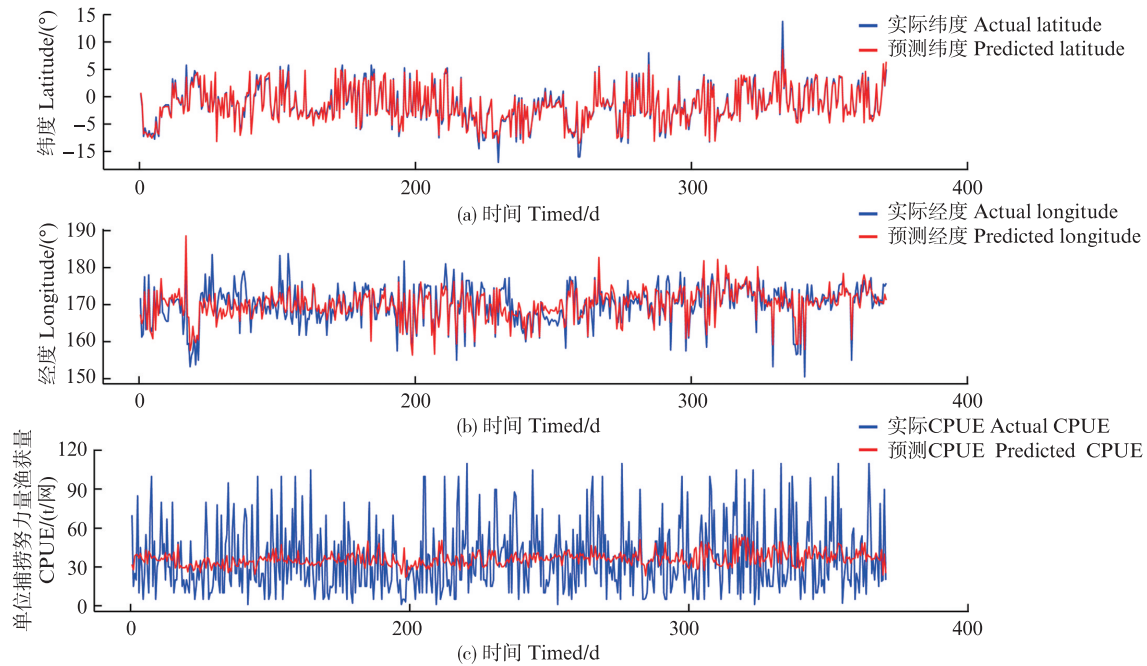
结果表明,预测的经纬度曲线波动基本与实际值相符,差异较小;预测的CPUE曲线波动幅度略小于实际值,但变化趋势与实际CPUE的时间序列变化基本一致。预测CPUE的最大值与最小值之差为31 t/网,而实际CPUE的差值为89 t/网。实际CPUE的平均值为36.39 t/网,预测CPUE的平均值为36.24 t/网,二者的均值十分接近,表明模型预测效果良好。

对LSTM模型预测的经纬度和CPUE与实际值进行相关性分析(图4),结果显示预测纬度与实际纬度之间的 R^2 为0.97,预测经度与实际经度

之间的 R^2 为0.65,而预测CPUE与实际CPUE之间的 R^2 为0.03。结果表明模型在经纬度预测方面具备较强的解释能力,但在CPUE的预测上表现相对有限。

2.5 鲷栖息地分布预测

本研究将2022年划分为4个季度:1—3月为第一季度,4—6月为第二季度,7—9月为第三季度,10—12月为第四季度,并利用LSTM模型预测鲷潜在栖息地分布区域,同时与实际渔场的空间分布进行叠加分析。分析结果显示,2022年鲷实际渔场的分布为12.25°S~13.75°N和150°E~170°W,主要集中在瑙鲁、基里巴斯(吉尔伯特群岛)和图瓦卢等3个专属经济区。第二季度和第三季度渔场分布较为分散,西延至巴布亚新几内亚专属经济区,但渔场仍主要集中于东部的3个经济区,其中基里巴斯(吉尔伯特群岛)渔场最为密集,种群丰度最高。



横轴表示距离基准时间2022年1月1日的天数。

The horizontal axis represents the number of days since the reference date, January 1, 2022.

图3 滞后10 d的LSTM模型预测结果和实际值的时序对比图

Fig. 3 Time series comparison between the predicted results of the LSTM model with a 10-day lag and the actual values

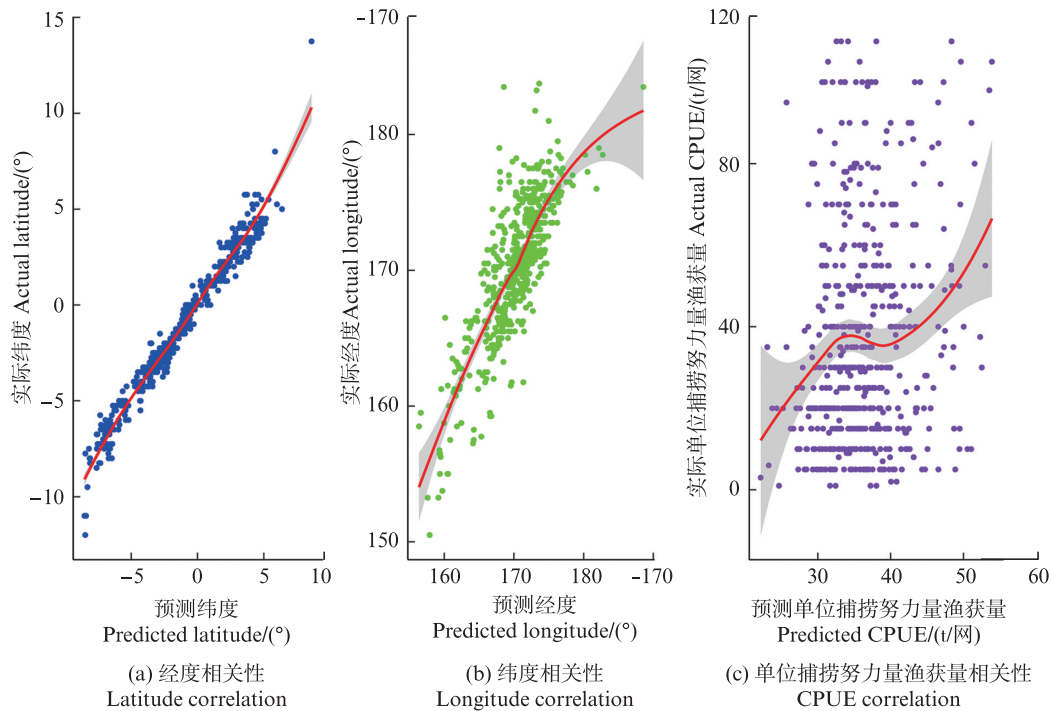


图4 滞后10 d的LSTM模型预测结果和实际值的相关性分析

Fig. 4 Correlation analysis between the predicted results of the LSTM model with a 10-day lag and the actual values

根据LSTM模型预测,2022年鳀栖息地的分布为 $12^{\circ}\text{S}\sim 10^{\circ}\text{N}$ 和 $155^{\circ}\text{E}\sim 175^{\circ}\text{W}$ 。第一、第二季度的潜在栖息地主要集中在基里巴斯专属经济区的东部边缘(图5和图6),第三季度分布较为

分散,并开始向西延伸至巴布亚新几内亚专属经济区(图7)。第四季度预测显示,渔场仍主要分布在瑙鲁、基里巴斯和图瓦卢等3个专属经济区(图8)。

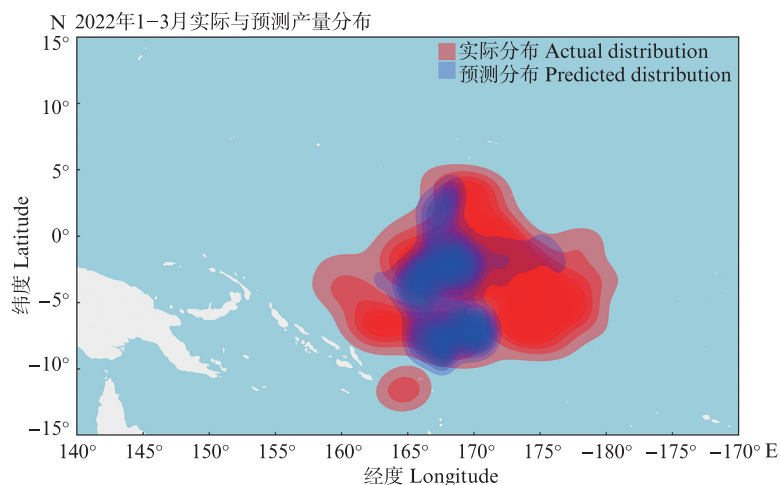


图5 2022年1—3月基于LSTM模型预测的鲹潜在栖息地和实际渔场分布对比图
Fig. 5 Comparison of distribution of potential habitats of skipjack and actual fishing grounds predicted by LSTM model from January to March 2022

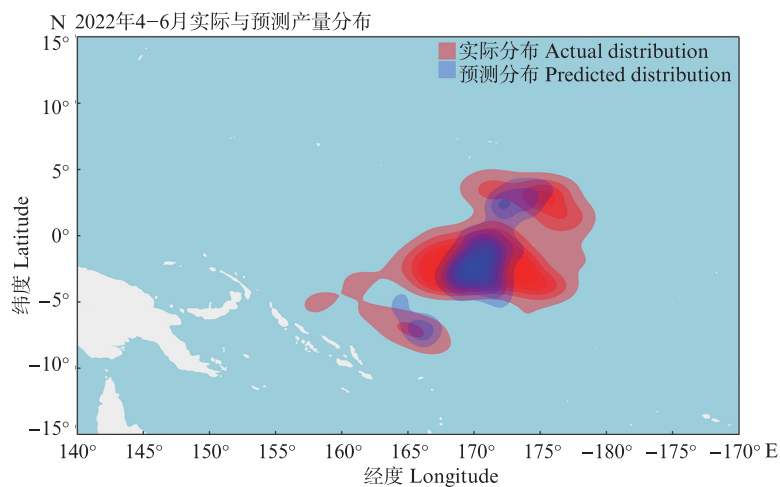


图6 2022年4—6月基于LSTM模型预测的鲹潜在栖息地和实际渔场分布对比图
Fig. 6 Comparison of distribution of skipjack potential habitat and actual fishing ground predicted by LSTM model from April to June 2022

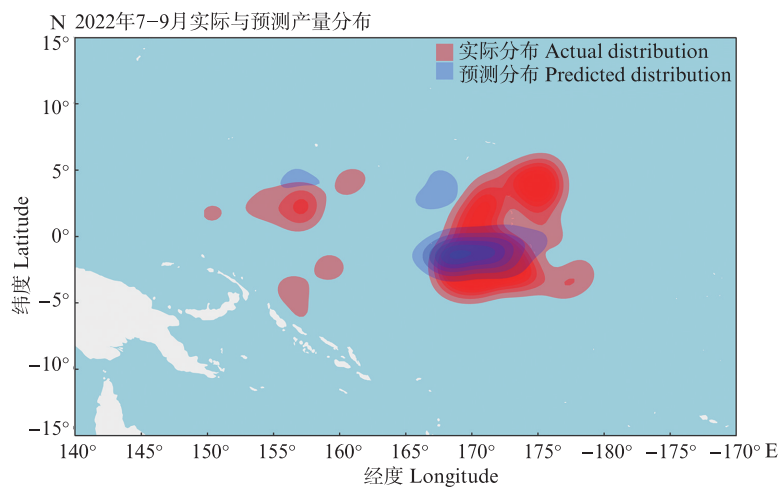


图7 2022年7—9月基于LSTM模型预测的鲹潜在栖息地和实际渔场分布对比图
Fig. 7 Comparison of distribution of potential habitats of skipjack and actual fishing grounds predicted by LSTM model from July to September 2022

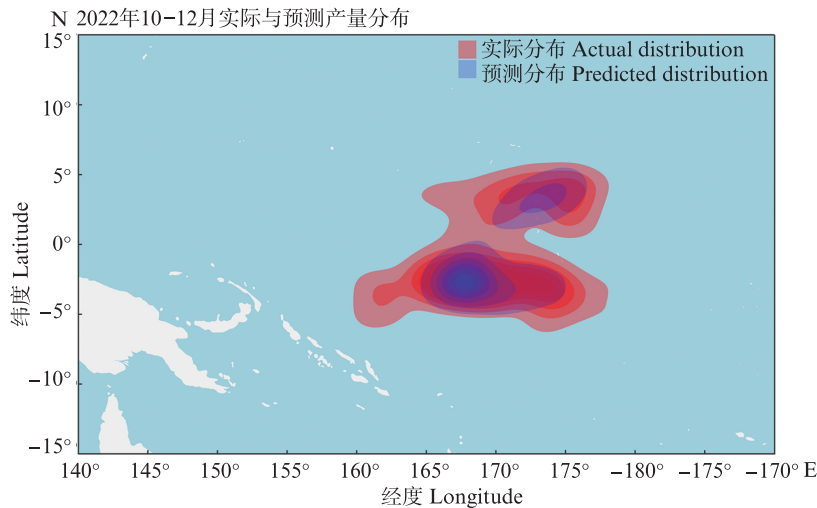


图8 2022年10—12月基于LSTM模型预测的鲣潜在栖息地和实际渔场分布对比图
Fig. 8 Comparison of distribution of potential habitats of skipjack and actual fishing grounds predicted by LSTM model from October to December 2022

综上所述,LSTM模型预测的栖息地空间分布与实际渔场分布较为吻合。预测结果表明,栖息地的分布较实际渔场更为集中,与高密度渔场的实际分布区域基本重叠,主要集中在赤道附近及150°E以东区域。如图5~图8,其中红色表示实际渔场分布,蓝色表示预测渔场分布。

3 讨论

LSTM模型能够有效克服传统静态模型的局限性,通过引入记忆单元,能够捕捉环境因子的动态变化和长期依赖关系,而传统模型通常仅关注单一时间点或短期数据,难以全面反映环境变化对鱼类栖息地分布的累计效应。此外,LSTM模型可以灵活建模多种滞后环境变量与渔场分布之间的复杂关系。通过综合不同的海洋环境因子和气象变量,该模型能够构建更加精确的渔场预测模型。在处理非线性时间序列数据方面,LSTM模型展现出显著优势^[38]。与基于经验假设的传统统计方法相比^[39-41],LSTM更加关注数据本身潜在模式的挖掘,减少了对主观假设的依赖。这种数据驱动的建模方法使其在滞后效应研究中表现出较高的客观性和适用性,能够更精确地揭示环境动态与渔业分布之间的复杂关系。

本研究表明,在 $0.25^{\circ} \times 0.25^{\circ}$ 分辨率下,滞后10 d的海洋环境因子对鲣CPUE时空分布的预测效果最佳,表明鲣空间分布受过去短期内环境累计效应的影响。与滞后15 d的模型相比,滞后10 d的模型预测精度更高,表明滞后10 d可

能是这些因子对鲣分布影响的一个最优时间窗口,既能反映环境条件对种群动态的积累效应,又反映了过长滞后时间会造成的影响衰减或随机性增强。如POWER^[42]通过种群动态模型分析美洲红点鲑(*Salvelinus fontinalis*)的栖息地变化模式发现,当环境条件变化时,物种栖息地分布的变化幅度会随着环境变化幅度的增加而增大。然而,当环境变化达到某一临界点后,栖息地分布的变化幅度开始逐渐减弱。这表明环境条件对种群动态的积累效应存在一定的阈值。

鲣对海洋环境变化的响应可能具有特定的时滞特征,10 d的滞后时间或反映其对环境变化的生理或行为调整周期。例如,从初级生产力的富集到高生产力区域的形成通常存在2至3周的滞后效应^[32]。鱼类可能因觅食需求而在特定海域周期性出现,从而在其空间分布上表现出显著的环境响应。这表明环境变化对鲣栖息地的影响不仅具有一定的滞后性,还反映了鱼类生态行为对资源分布的依赖关系。

尽管滞后时间为10 d的模型在MSE和MAE指标上表现优于其他滞后时间构建的模型,但其优势并不显著。导致该结果的原因可能包括:本研究的时间跨度仅为4年,样本量相对有限,可能限制了模型的泛化能力;神经元数量的组合未能通过随机搜索充分优化,可能导致模型训练的效果尚不完全;所有模型均基于 $0.25^{\circ} \times 0.25^{\circ}$ 的空间分辨率进行分析,未能考虑不同空间尺度下滞后时间的最优性;针对多种环境变量在不同空间尺

度上的综合分析及其最佳滞后时间的研究仍较为缺乏;未来研究可以尝试采用更精细的时间间隔进行分析,如以天为单位构建滞后0、1、2、…、15 d的环境数据构建LSTM模型,比较不同滞后时间下模型的预测性能,进一步探讨滞后时间对预测性能的影响。

基于滞后10 d环境因子构建的LSTM模型为提升鲣渔场预测的科学性提供了新的技术支持。该模型可帮助渔民优化作业区域选择,通过监测诸如温跃层深度和营养盐浓度等关键环境变量的动态变化,提前判断适宜的捕捞区域,从而减少无效捕捞的时间和资源消耗。此外,渔业管理部门可以利用该模型监控海洋环境变化趋势,例如海水升温或酸化,从而科学制定捕捞计划,缓解环境变化可能对渔业资源可持续性带来的潜在影响。

作者声明本文无利益冲突。

参考文献:

- [1] SUND P N, BLACKBURN M, WILLIAMS F. Tunas and their environment in the Pacific Ocean: a review[J]. *Oceanography and Marine Biology: An annual review*, 1981, 19: 443-512.
- [2] ARRIZABALAGA H, DUFOUR F, KELL L, et al. Global habitat preferences of commercially valuable tuna [J]. *Deep Sea Research Part II: Topical Studies in Oceanography*, 2015, 113: 102-112.
- [3] 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.
- [4] 范江涛. 南太平洋长鳍金枪鱼延绳钓渔业渔情预报研究[D]. 上海: 上海海洋大学, 2011.
FAN J T. Forecasting of fishing ground for Albacore (*Thunnus alalunga*) in the South Pacific [D]. Shanghai: Shanghai Ocean University, 2011.
- [5] 宋函欣. 我国中西太平洋金枪鱼围网渔业发展研究 [D]. 上海: 上海海洋大学, 2022.
SONG H X. Study on the development of China tuna purse seinefishery in the Western and Central Pacific Ocean [D]. Shanghai: Shanghai Ocean University, 2022.
- [6] 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.
- [7] 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.
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.
- [8] 徐国强, 朱文斌, 张洪亮, 等. 基于GAM模型分析印度洋大眼金枪鱼和黄鳍金枪鱼渔场分布与不同环境因子关系[J]. *海洋学报*, 2018, 40(12): 68-80.
XU G Q, ZHU W B, ZHANG H L, et al. Relationship between fishing grounds of *Thunnus obesus* and *Thunnus albacores* with environmental factors in the Indian Ocean based on generalized additive model [J]. *Haiyang Xuebao*, 2018, 40(12): 68-80.
- [9] 周为峰, 黎安舟, 纪世建, 等. 基于贝叶斯分类器的南海黄鳍金枪鱼渔场预报模型[J]. *海洋湖沼通报*, 2018 (1): 116-122.
ZHOU W F, LI A Z, JI S J, et al. Forecasting model for yellowfin tuna (*Thunnus albacares*) fishing ground in the South China Sea based on Bayes classifier [J]. *Transactions of Oceanology and Limnology*, 2018 (1): 116-122.
- [10] 张聪, 周为峰, 唐峰华, 等. 基于机器学习的中西太平洋黄鳍金枪鱼渔场预报模型[J]. *农业工程学报*, 2022, 38(15): 330-338.
ZHANG C, ZHOU W F, TANG F H, et al. Forecasting models for yellowfin tuna fishing ground in the central and western Pacific based on machine learning [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022, 38(15): 330-338.
- [11] YEN K W, WANG G H, LU H J. Evaluating habitat suitability and relative abundance of skipjack (*Katsuwonus pelamis*) in the Western and Central Pacific during various El Niño events [J]. *Ocean & Coastal Management*, 2017, 139: 153-160.
- [12] POMEROY R, PHANG K H W, RAMDASS K, et al. Moving towards an ecosystem approach to fisheries management in the Coral Triangle region [J]. *Marine Policy*, 2015, 51: 211-219.
- [13] ARTETXE-ARRATE I, FRAILE I, MARSAC F, et al. A review of the fisheries, life history and stock structure of tropical tuna (skipjack *Katsuwonus pelamis*, yellowfin *Thunnus albacares* and bigeye *Thunnus obesus*) in the Indian Ocean [J]. *Advances in Marine Biology*, 2021, 88: 39-89.
- [14] 张健, 宋厚成, 刘文俊, 等. 基于LSTM的气候变化对中西太平洋围网黄鳍金枪鱼和鲣CPUE影响的滞后效应 [J]. *海洋学报*, 2024, 46(7): 62-72.
ZHANG J, SONG H C, LIU W J, et al. Lag effect of

- climate change on CPUE of *Thunnus albacares* and *Katsuwonus pelamis* in the western and central Pacific Ocean purse seine fishery: an LSTM-based study [J]. Haiyang Xuebao, 2024, 46(7): 62-72.
- [15] 杜艳玲, 马玉玲, 汪金涛, 等. 基于ConvLSTM-CNN预测太平洋长鳍金枪鱼时空分布趋势[J]. 海洋通报, 2024, 43(2): 174-187.
- DU Y L, MA Y L, WANG J T, et al. Prediction of spatial-temporal distribution trend of Pacific albacore tuna based on ConvLSTM-CNN[J]. Marine Science Bulletin, 2024, 43(2): 174-187.
- [16] DRUON J N, CHASSOT E, MURUA H, et al. Skipjack tuna availability for purse seine fisheries is driven by suitable feeding habitat dynamics in the Atlantic and Indian Oceans[J]. Frontiers in Marine Science, 2017, 4: 315.
- [17] HSU T Y, CHANG Y, LEE M A, et al. Predicting skipjack tuna fishing grounds in the Western and Central Pacific Ocean based on high-spatial-temporal-resolution satellite data[J]. Remote Sensing, 2021, 13(5): 861.
- [18] ZAINUDDIN M, FARHUM A, SAFRUDDIN S, et al. Detection of pelagic habitat hotspots for skipjack tuna in the Gulf of Bone-Flores Sea, southwestern Coral Triangle tuna, Indonesia [J]. PLoS One, 2017, 12 (10) : e0185601.
- [19] MATSUMOTO T, SATOH K, SEMBA Y, et al. Comparison of the behavior of skipjack (*Katsuwonus pelamis*), yellowfin (*Thunnus albacares*) and bigeye (*T. obesus*) tuna associated with drifting FADs in the equatorial central Pacific Ocean [J]. Fisheries Oceanography, 2016, 25(6): 565-581.
- [20] BERNAL D, BRILL R W, DICKSON K A, et al. Sharing the water column: physiological mechanisms underlying species-specific habitat use in tunas [J]. Reviews in Fish Biology and Fisheries, 2017, 27 (4) : 843-880.
- [21] MUGO R, SAITOH S I, NIHIRA A, et al. Habitat characteristics of skipjack tuna (*Katsuwonus pelamis*) in the western North Pacific: a remote sensing perspective [J]. Fisheries Oceanography, 2010, 19(5): 382-396.
- [22] SCHAEFER K M, FULLER D W. Vertical movement patterns of skipjack tuna (*Katsuwonus pelamis*) in the eastern equatorial Pacific Ocean, as revealed with archival tags[J]. Fishery Bulletin, 2007, 105(3): 379-389.
- [23] KERR S R, RYDER R A. Current approaches to multispecies analyses of marine fisheries [J]. Canadian Journal of Fisheries and Aquatic Sciences, 1989, 46(3): 528-534.
- [24] HOERL A E, KENNARD R W. Ridge regression: biased estimation for nonorthogonal problems [J]. Technometrics, 1970, 12(1): 55-67.
- [25] TIBSHIRANI R. Regression shrinkage and selection via the lasso [J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 1996, 58 (1) : 267-288.
- [26] ZOU H, HASTIE T. Addendum: regularization and variable selection via the elastic net [J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2005, 67(5): 768.
- [27] TAO D P, LIN X, JIN L W, et al. Principal component 2-D long short-term memory for font recognition on single Chinese characters [J]. IEEE Transactions on Cybernetics, 2016, 46(3): 756-765.
- [28] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9 (8) : 1735-1780.
- [29] GERS F A, SCHMIDHUBER J, CUMMINS F. Learning to forget: continual prediction with LSTM [J]. Neural Computation, 2000, 12(10): 2451-2471.
- [30] 孙强, 薛存金, 刘敬一, 等. 全球海洋初级生产力与海洋环境要素时空关联模式挖掘分析[J]. 海洋环境科学, 2020, 39(3): 340-347, 352.
- SUN Q, XUE C J, LIU J Y, et al. Spatiotemporal association patterns between marine net primary production and environmental parameters in a view of data mining [J]. Marine Environmental Science, 2020, 39 (3): 340-347, 352.
- [31] 李文权, 王宪, 李淑英. 光照强度及环境要素对海洋初级生产力的影响[J]. 厦门大学学报(自然科学版), 1989, 28(4): 423-426.
- LI W Q, WANG X, LI S Y. Effects of light intensity and environment factors on marine primary productivity [J]. Journal of Xiamen University (Natural Science), 1989, 28(4): 423-426.
- [32] FONTENEAU A, LUCAS V, TEWKAI E, et al. Mesoscale exploitation of a major tuna concentration in the Indian Ocean [J]. Aquatic Living Resources, 2008, 21(2): 109-121.
- [33] CHAI T, DRAXLER R R. Root mean square error (RMSE) or mean absolute error (MAE)? - arguments against avoiding RMSE in the literature [J]. Geoscientific Model Development, 2014, 7(3): 1247-1250.
- [34] DE MYTTENAERE A, GOLDEN B, LE GRAND B, et al. Mean absolute percentage error for regression models [J]. Neurocomputing, 2016, 192: 38-48.
- [35] FOSTER R G, ROENNEBERG T. Human responses to the geophysical daily, annual and lunar cycles [J]. Current Biology, 2008, 18(17): R784-R794.
- [36] HODSON T O. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not [J]. Geoscientific Model Development, 2022, 15 (14) :

- 5481-5487.
- [37] NALLURI M, PENTELA M, ELURI N R. A scalable tree boosting system: XG boost[J]. International Journal of Research Studies in Science, Engineering and Technology, 2020, 7(12): 36-51.
- [38] CAVIESES-NÚÑEZ R, OJEDA-RUIZ M A, FLORES-IRIGOLLEN A, et al. Focused small-scale fisheries as complex systems using deep learning models [J]. Latin American Journal of Aquatic Research, 2021, 49 (2) : 342-353.
- [39] BENAVIDES I F, ROMERO-LEITON J P, SANTACRUZ M, et al. Applying seasonal time series modeling to forecast marine fishery landings for six species in the Colombian Pacific Ocean[J]. Regional Studies in Marine Science, 2022, 56: 102716.
- [40] LEEB, SUHARTONO S, LIMA, et al. Forecasting world tuna catches with ARIMA-spline and ARIMA-neural networks models[J]. Walailak Journal of Science and Technology (WJST), 2021, 18(17): 9726.
- [41] WU Y L, LAN K W, TIAN Y J. Determining the effect of multiscale climate indices on the global yellowfin tuna (*Thunnus albacares*) population using a time series analysis[J]. Deep Sea Research Part II : Topical Studies in Oceanography, 2020, 175: 104808.
- [42] POWER M. Assessing the effects of environmental stressors on fish populations [J]. Aquatic Toxicology, 1997, 39(2): 151-169.

Habitat prediction of skipjack in the Western and Central Pacific based on LSTM model

ZHOU Cheng^{1,2,3,4}, ZHOU Xiang¹, HU Yuanyuan¹, LIU Liwen¹

(1. College of Marine Living Resource Sciences and Management, Shanghai Ocean University, Shanghai 201306, China; 2. National Engineering Research Center for Oceanic Fisheries, Shanghai 201306, China; 3. Key Laboratory of Sustainable Exploitation of Oceanic Fisheries Resources, Ministry of Education, Shanghai 201306, China; 4. Key Laboratory of Sustainable Utilization of Oceanic Fisheries, Ministry of Agriculture and Rural Affairs, Shanghai 201306, China)

Abstract: To address the limitations of traditional habitat prediction models in capturing the lagged effects of environmental factors with time series information on tuna spatial distribution, this study utilized tuna purse-seine fishing log data from 2021 to 2024. Long-short term memory (LSTM) neural network models were constructed with lag durations of 1 day, 5 days, 10 days, and 15 days to predict catch per unit of effort (CPUE) and geographic coordinates (latitude and longitude). The findings indicate that the 10-day lag model exhibited the highest accuracy, with a mean square error (MSE) of 0.018 7 and a mean absolute error (MAE) of 0.077 6, suggesting that the spatial distribution of skipjack is influenced by cumulative short-term environmental effects. Validation of the optimal model demonstrated the R^2 of 0.97 for predicted versus actual latitude and 0.65 for longitude, indicating a strong alignment between predicted and observed spatial distributions. This research offers new insights into the dynamic mechanisms underlying skipjack tuna habitat characteristics and ecological processes. Furthermore, it provides critical references for the scientific management of skipjack purse seine fisheries in the Western and Central Pacific Ocean.

Key words: *Katsuwonus pelamis*; habitat prediction; LSTM model; the Western and Central Pacific