

DOI: 10.12264/JFSC2025-0024

基于 LASSO 回归方法的南太平洋长鳍金枪鱼补充量预测

王扬^{1, 2}, 朱江峰^{1, 2, 3}, 张帆^{1, 2, 3}, 耿喆^{1, 2*}

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, 农业农村部大洋渔业可持续利用重点实验室, 上海 201306;
3. 上海海洋大学, 国家远洋渔业工程技术研究中心, 上海 201306

摘要: 种群补充是维持资源可持续的关键生物学过程, 准确理解和预测补充量的变异性已经成为渔业管理的核心问题。以往对南太平洋长鳍金枪鱼(*Thunnus alalunga*)补充量的预测方法在处理环境变量间多重共线性及筛选关键驱动变量方面时存在显著局限性, 容易导致模型过拟合及预测精度下降。为弥补以上不足, 本研究引入 LASSO 回归算法以优化预测变量筛选过程并提升预测精度。基于 2000—2017 年观测数据构建模型, 并结合 CMIP6 多模式气候预测数据, 对 2018—2100 年补充量进行长期预测。结果表明, LASSO 方法通过收缩无关变量, 有效消除变量冗余, 提高了模型预测精度。最佳模型解释率为 45.9%, 其中海表温度(SST)和混合层深度(MLD)为预测补充量的重要因子。预测结果显示, 在二氧化碳较高排放情形下(SSP585, SSP370), 在 2070 年代种群补充量趋近于零, 种群崩溃风险显著升高; 在低碳路径(SSP126)下, 补充量仍呈现持续性衰减趋势。本研究结果为渔业资源预测及变量选择介绍了有效方法, 并构建了气候-补充耦合预测模型, 为制定适应性管理策略、规避种群崩溃风险提供了量化决策支持。

关键词: 补充量预测; LASSO 方法; CMIP6; 长鳍金枪鱼

中图分类号: S931

文献标志码: A

文章编号: 1005-8737-(2025)07-0914-10

鱼类种群补充过程指新生个体成功加入种群并达到性成熟的生物学过程, 该过程不仅决定了种群生产力及潜在渔获量, 对过度开发种群的恢复也具有关键作用, 是维持资源可持续的生物学基础^[1-2]。渔业资源丰度的时空变异很大程度上归因于自然或人为因素引起的补充量变化。由于种间关系、种内竞争等生物因素和温度、洋流等非生物因素的相互作用, 补充过程常呈现出年间变异性, 气候变化造成的长期海洋生态变化进一步加剧了不确定性, 理解和预测补充量的变异性已经成为渔业管理的核心问题^[3]。

长鳍金枪鱼(*Thunnus alalunga*)是具有重要经济价值的目标鱼种, 广泛分布于温带海域^[4]。在太

平洋, 该鱼种以赤道为界分为南、北 2 个种群^[5], 其中位于中西太平洋渔业委员会(Western and Central Pacific Fisheries Commission, WCPFC)管辖范围内的南太平洋长鳍金枪鱼捕捞量占整个中西太长鳍金枪鱼总产量的 70%^[6]。在长鳍金枪鱼种群动态研究中, 通常将表征亲体量对补充量影响的参数陡度(steeptness, h)设定值为 0.8, 意味着亲体量变化对其补充量影响较小, 种群产卵量相对稳定, 补充量变化主要由环境驱动^[7]。由于长鳍金枪鱼生长速度较慢, 在气候变化加剧的背景下, 其幼鱼生长至补充年龄的过程更易受到环境影响, 从而影响最终补充量^[8]。根据以往研究发现, 长鳍金枪鱼的补充过程受饵料丰度、海表温度、海平面

收稿日期: 2025-02-07; 修订日期: 2025-04-09.

基金项目: 国家重点研发计划项目(2024YFD2400602); 农业农村部 2024 年度全球重要鱼种资源动态监测评估项目(D-8025-24-5001).

作者简介: 王扬(1993–), 女, 博士后, 从事气候变化与鱼类种群动态研究. E-mail: yan-wang@shou.edu.cn

通信作者: 耿喆, 讲师, 从事中西太平洋渔业资源评估及管理. E-mail: zgeng@shou.edu.cn

高度等环境参数调控^[9-11], 在极端气候变化发生年份, 补充量表现出显著增加或减少^[12]。因此, 通过建立气候变量与种群补充之间预测模型, 能更有效地评估气候变化对长鳍金枪鱼资源动态的影响, 为渔业管理提供参考。

基于气候模型的补充量预测技术虽发展迅速, 但在实际渔业管理中的应用仍较为有限, 主要原因在于预测精度较低, 存在偏差较高^[13]。以往在构建补充量预测模型时, 常采用时间序列分析、广义加性模型、贝叶斯模型等方法^[14-16]。随着计算机技术不断发展, 人工智能技术也逐渐应用于渔业管理当中构建预测模型^[17-18]。然而, 这些模型在处理多个预测变量时, 由 2 个或多个自变量高度相关引发的共线性问题可能导致模型回归系数不稳定、参数解释性下降、模型显著性降低等情况, 从而影响预测精度^[19]。此外, 变量选择常通过构建包含所有可能变量组合的模型并使用特定指标(如 AIC 值)进行比较, 进而筛选最佳模型, 但当预测变量较多时, 模型复杂性及运行时间显著增加, 可能导致参数估算偏差、过拟合或未充分利用模型空间等问题^[20-21]。本研究引入了基于正则化(Regularization)技术的最小绝对收缩和选择算子回归分析(Least Absolute Shrinkage and Selection Operator, LASSO)^[22]以解决模型预测中遇到的上述问题, 提高预测精度。与常见变量选择方法不同, 通过在模型中添加收缩项对回归系数施加约束和惩罚, 将不重要变量的回归系数压缩为 0, 实现自动选择重点变量及解决变量间共线性问题, 相比逐步回归方法其预测误差降低^[23]。LASSO 回归在医学、工程等领域应用较多, 应用于渔业领域的研究还十分有限^[24]。

本研究使用 LASSO 回归分析识别显著影响南太平洋长鳍金枪鱼补充量的环境变量, 并基于政府间气候变化专门委员会(Intergovernmental Panel on Climate Change, IPCC)耦合模式对比计划第 6 阶段(Coupled Model Intercomparison Project Phase 6, CMIP6)不同模式下气候预测数据对补充量进行长期预测。研究结果有助于提高对气候变化影响下长鳍金枪鱼资源动态的评估精度, 增强

极端气候事件对补充量潜在影响的预测能力, 并为科学制定渔业管理措施提供重要参考。

1 材料与方法

1.1 环境因子及补充量

由于高度洄游物种难以实现生物学监测, 长鳍金枪鱼长期补充量数据由 WCPFC 采纳的 2021 年南太平洋长鳍金枪鱼资源评估拟合结果中提取, 时间从 1960—2019 年, 以季度为单位^[7]。评估模型中虽对研究区域进行了分区, 但在估算补充量时以每年同比例划分给各个区域, 难以体现准确的生物学意义, 因此本研究采用的补充量为每季度南太平洋海域的总和。由于评估结果中输出的补充量仅到 2017 年, 此后为预测值, 因此本研究选择的补充量时间序列为 1960—2017 年。

基于以往关于环境对长鳍金枪鱼影响研究^[5,9,25], 选取预计对长鳍金枪鱼补充有重要影响的 5 个预测因子: 海表面温度(sea surface temperature, SST)、叶绿素浓度(chlorophyll, Chl)、海表面盐度(sea surface salinity, SSS)、海表面高度(sea surface height, SSH)和混合层深度(mixed layer depth, MLD)。环境变量分为历史观测和基于 CMIP6 预测的未来数据(表 1)。

历史观测数据 SST 来自美国国家海洋和大气管理局(National Oceanic and Atmospheric Administration, NOAA)重构数据(<https://psl.noaa.gov/data/gridded/data.noaa.ersst.v5.html>), 时间为 1960—2019 年, 分辨率为月, 空间分辨率为 $2^{\circ} \times 2^{\circ}$ ^[26]。Chl 来自美国国家航天局(National Aeronautics and Space Administration, NASA)的遥感观测数据(<https://oceandata.sci.gsfc.nasa.gov/directdataaccess/>), 时间从 2000—2019 年, 时间分辨率为天, 空间分辨率为 9 km ^[27]。SSS、SSH 及 MLD 均来自欧洲中期天气预报中心(European Centre for Medium-Range Weather Forecasts, ECMWF), 时间从 1960—2019 年, 分辨率为月, 空间分辨率为 $0.25^{\circ} \times 0.25^{\circ}$ (<https://cds.climate.copernicus.eu/datasets/reanalysis-oras5?tab=overview>)^[28]。

CMIP6 结合了社会经济发展的新型共享路径(shared socio-economic pathways, SSPs), 根据未

来二氧化碳不同的排放趋势设置预测情形, 本研究根据强弱从中选取 SSP126、SSP370 及 SSP585 情形对补充量进行预测, 时间跨度为 2015—2100 年, 气候模式为 CESM2-WACCM, 时间分辨率

为月, 空间分辨率为 $1^{\circ}\times 1^{\circ}$ (<https://aims2.llnl.gov/search/cmip6/>)。为和补充量数据匹配, 将上述环境数据以季度为单位, 计算整个研究区域平均值。

表 1 环境变量历史和预测数据来源、空间分辨率、时间及时间分辨率

Tab. 1 Sources, spatial resolution, time period, and temporal resolution of historical and forecast environmental variables

预测因子 redictors	历史观测 historically observed				预测 projected				
	模式代码 model code	空间分辨率 spatial resolution	时间跨度 time duration	时间分辨率 temporal resolution	模式代码 model code	模式输出代码 output code	空间分辨率 spatial resolution	时间跨度 time duration	时间分辨率 temporal resolution
海表面温度 SST	ERSST	2°	1960–2019	月 monthly	CESM2-WACCM	tos	1°	2015–2100	月 monthly
叶绿素浓度 Chl	MODIS	9 km	2000–2019	天 daily	CESM2-WACCM	chl	1°	2015–2100	月 monthly
海表面盐度 SSS	ECMWF	0.25°	1960–2019	月 monthly	CESM2-WACCM	so	1°	2015–2100	月 monthly
海表面高度 SSH	ECMWF	0.25°	1960–2019	月 monthly	CESM2-WACCM	zos	1°	2015–2100	月 monthly
混合层深度 MLD	ECMWF	0.25°	1960–2019	月 monthly	CESM2-WACCM	mldst	1°	2015–2100	月 monthly

1.2 LASSO 回归变量选择过程

均方误差(mean square error, MSE)是衡量模型预测值与真实值之间差异的指标, MSE 越小, 说明回归模型预测性能越好。MSE 计算公式如下:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - y_i^*)^2 \tag{1}$$

其中, N 为样本总数, y_i 为第 i 个观测值(真实值), y_i^* 为第 i 个预测值。

LASSO 在损失函数(loss function)中引入正则化项, 对回归系数施加收缩约束, 将无关变量回归系数收缩为零。被收缩为零的变量即被剔除, 从而实现变量筛选, 达到最小化 MSE, 公式如下:

$$LASSO\ MSE = \min_{\beta_0, \beta} \frac{1}{N} \sum_{i=1}^N (y_i - \beta_0 - x_i^T \beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \tag{2}$$

其中, N 为观测值数量, y_i 为第 i 个观测值的响应变量, x_i 为长度为 p 的预测变量向量, β_0 为常数项, β_j 为第 j 个预测变量的回归系数, λ 为惩罚系数, 决定惩罚力度。对应最低 MSE 的 λ 值为最优 λ 解。对于给定 λ , LASSO 解出每个变量的回归系数 β_j 。当 β_j 为 0 时, 表示剔除该变量, β_j 不为 0 表示保留该变量。

1.3 模型构建

长鳍金枪鱼成鱼常分布于南纬 30° 以北区域^[29], 随着洄游路线到特定区域觅食和产卵, 通常在南太

平洋南纬 5° 至 25° 的热带和亚热带海域表层海域产卵, 产卵高峰期为 10 月至 12 月^[30]。产卵后的幼鱼到其 1 岁龄时, 开始向南迁移进入亚热带辐合带, 分布范围为南纬 25° 至 50° 、东经 140° 至西经 130° 海域, 定义其为补充区域^[31]。根据上述研究, 将研究区域划分为产卵区和补充区域(图 1)。

根据上述定义的补充区和产卵区, 本研究设置了 5 个模型以探讨不同区域的环境数据对预测的结果的影响(表 2), 基于均方根误差(root mean square error, RMSE)和模型解释率(r-squared)进行模型筛选。基于关键期假说及金枪鱼生物学研究^[32], 金枪鱼产卵后的幼鱼食物具有高度选择性, 依赖于有限种类浮游动物, 因此代表食物指标的 Chl 是仔稚鱼存活和补充量关键因素^[33]。由于 LASSO 模型无法处理连续缺失值, Chl 仅可获取 2000 年后数据, 因此模型情形包含了整个时间跨度及 2000 年以后时间跨度。模型构建如下:

按时间顺序, 取 80% 数据作为训练集, 20% 数据作为测试集。模型预测分为 2 个阶段: (1) 2000—2017 年, 基于观测环境数据预测, 进行模型验证; (2) 2018—2100 年, 结合 GCM 环境预测数据对未来补充量进行预测。最后使用相同数据和假设, 对基于 LASSO 回归的预测模型与传统最小二乘法(ordinary least squares, OLS)多元线性回归模型进行预测性能对比。

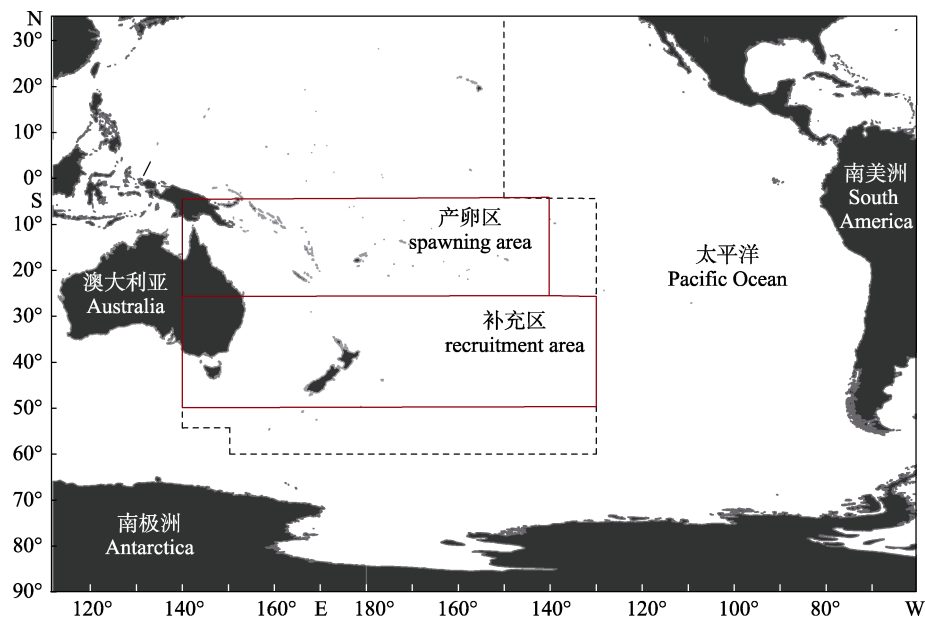


图 1 南太平洋长鳍金枪鱼产卵区及补充区分布
虚线内为 WCPFC 管辖区域.

Fig. 1 Spawning and recruitment areas of *Thunnus alalunga* in the South Pacific
Area within the dashed lines indicates the jurisdictional area of the WCPFC.

表 2 南太平洋长鳍金枪鱼补充量预测模型构建

Tab. 2 Details of recruitment prediction model construction for <i>Thunnus alalunga</i> in the South Pacific							
模型代码	model code	时间长度	time duration	补充量区域	recruitment area	环境变量区域	environmental variables area
	M1	1960–2017		南太平洋	South Pacific Ocean	南太平洋	South Pacific Ocean
	M2	2000–2017		南太平洋	South Pacific Ocean	南太平洋	South Pacific Ocean
	M3	2000–2017		南太平洋	South Pacific Ocean	产卵区	spawning area
	M4	2000–2017		南太平洋	South Pacific Ocean	补充区	recruitment area
	M5	2000–2017		南太平洋	South Pacific Ocean	产卵区及补充区	spawning and recruitment area

2 结果与分析

2.1 模型结果

模型分析结果见表 3。其中 M1 模型为包含整个序列数据, 模型解释方差较低仅为 0.108, 均方根误差较大为 2.568, 说明该模型对数据解释能力有限, 不适合作为预测模型。最优模型为 M2, 其解释方差最高为 0.459, RMSE 最低为 1.153。M2 保留了 5 个变量, 其中 SST、MLD、SSH 回归系数为负数, 与因变量呈负相关关系, 具体系数依次为-0.83、-0.73、-0.25, SSS 和 CHL 的回归系数为正数, 分别为 0.15 和 0.13。其余 3 组模型, 根据模型表现程度排名为 M4、M3 和 M5, 对应的解释方差分别为 0.428、0.403、0.391, 均方根误差分别为 1.292、1.524、1.804。

从变量筛选来看, SST、MLD、SSS、CHL 为所有模型均保留的变量。其中, SST 回归系数均为负数, 数值范围在-0.83 至-0.56 之间; MLD 回归系数均为负值为-0.73 至-0.62 之间; CHL 回归系数为正数, 在 M3 中回归系数最高为 0.26, 其他模型中均约为 0.1; SSS 回归系数为正值, 相对较低, 在 0.11~0.15 之间。SSH 仅在 M2 和 M4 模型中为保留变量, 该变量的回归系数分别为-0.25 和-0.16。

2.2 模型评估

交叉验证用于评估 LASSO 模型的泛化能力 (图 2), 可用来选择最佳正则化参数 λ 。图中左侧虚线为最小 MSE 对应 λ 值, 取该 λ 值的模型偏差最小, 拟合效果最佳。右侧虚线为误差在最小的

表 3 最终 LASSO 模型保留变量、变量系数，
及模型最终均方根误差和解释方差
Tab. 3 Retained variables, variable coefficients, RMSE,
and r-squared of the final LASSO models

模型代码 model code	保留变量 effective variables	系数 coefficient	均方根误差 RMSE	解释方差 r-squared
M1	海表面盐度 SSS	0.09	2.568	0.108
	海表面温度 SST	-0.04		
M2	海表面温度 SST	-0.83	1.153	0.459
	混合层深度 MLD	-0.73		
	海表面高度 SSH	-0.25		
	海表面盐度 SSS	0.15		
	叶绿素浓度 CHL	0.13		
M3	混合层深度 MLD	-0.62	1.524	0.403
	海表面温度 SST	-0.56		
	叶绿素浓度 CHL	0.26		
	海表面盐度 SSS	0.11		
	海表面高度 SSH	-0.16		
M4	混合层深度 MLD	-0.65	1.292	0.428
	海表面温度 SST	-0.63		
	海表面高度 SSH	-0.16		
	海表面盐度 SSS	0.12		
	叶绿素浓度 CHL	0.11		
M5	混合层深度 MLD	-0.68	1.804	0.391
	海表面温度 SST	-0.59		
	海表面盐度 SSS	0.15		
	叶绿素浓度 CHL	0.14		

一个标准方差范围内的 λ 值，在此 λ 值下，模型的拟合效果较好，但选取变量数相对较少，本研究以左侧虚线作为最佳 λ 值选项，图中顶部数据为保留变量个数。红色点代表每个模型在不同 λ 值下的平均 MSE，灰色误差条表示 MSE 的标准差。图中结果显示保留变量个数与模型筛选变量个数一致。随着 λ 值增大，所有模型的 MSE 均表现出逐渐上升的趋势，即随着正则化力度增强，模型误差增加。

图 3 为变量重要性得分，LASSO 将不重要变量系数直接缩小为 0。根据图中结果显示，SST 和 MLD 为最重要的 2 个变量；CHL 在 M3 中影响较为显著；SSS 虽保留在各个模型中，但重要性得分均较低；SSH 在 M3 和 M5 中系数直接缩减为 0，在 M2 和 M4 中影响效果也较小。

2.3 预测结果

图 4 为基于 4 个模型对补充量进行预测结果，预测分为 2 个阶段，第一阶段基于 2000—2017 年观测环境数据对模型进行验证；第二阶段基于 CMIP6 预测环境数据对 2018—2100 年补充量进行长期预测。

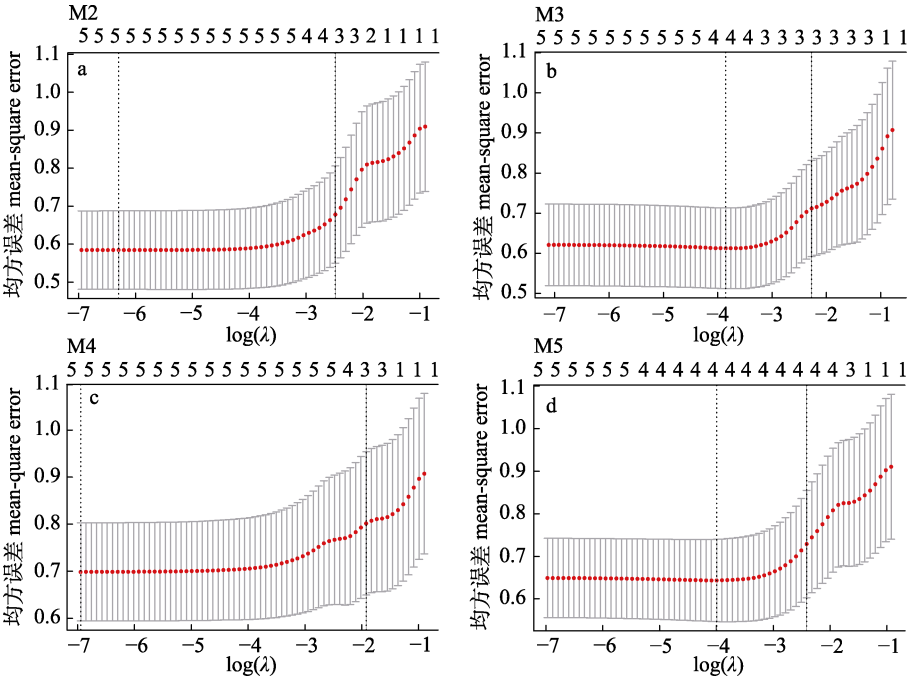


图 2 模型 M2 (a)、M3 (b)、M4 (c)、M5 (d)交叉检测结果
图中左侧虚线为最小 MSE 对应的 λ 值，右侧虚线为误差在最小的一个标准方差范围内的 λ 值。
Fig. 2 Cross-validation results of M2 (a), M3 (b), M4 (c), and M5 (d)
The left dashed line represents the λ value corresponding to the minimum MSE, and the right dashed line represents the λ value within one standard error range of the minimum error.

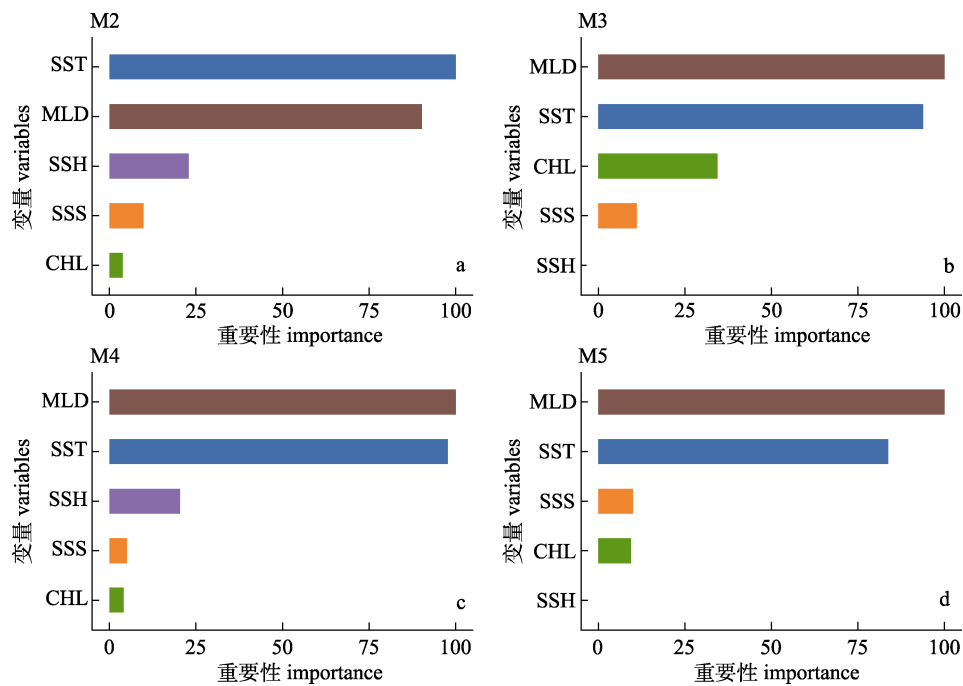


图 3 模型 M2 (a)、M3 (b)、M4 (c)、M5 (d) 中变量重要性得分

Fig. 3 Variable importance scores for M2 (a), M3 (b), M4 (c), and M5 (d)

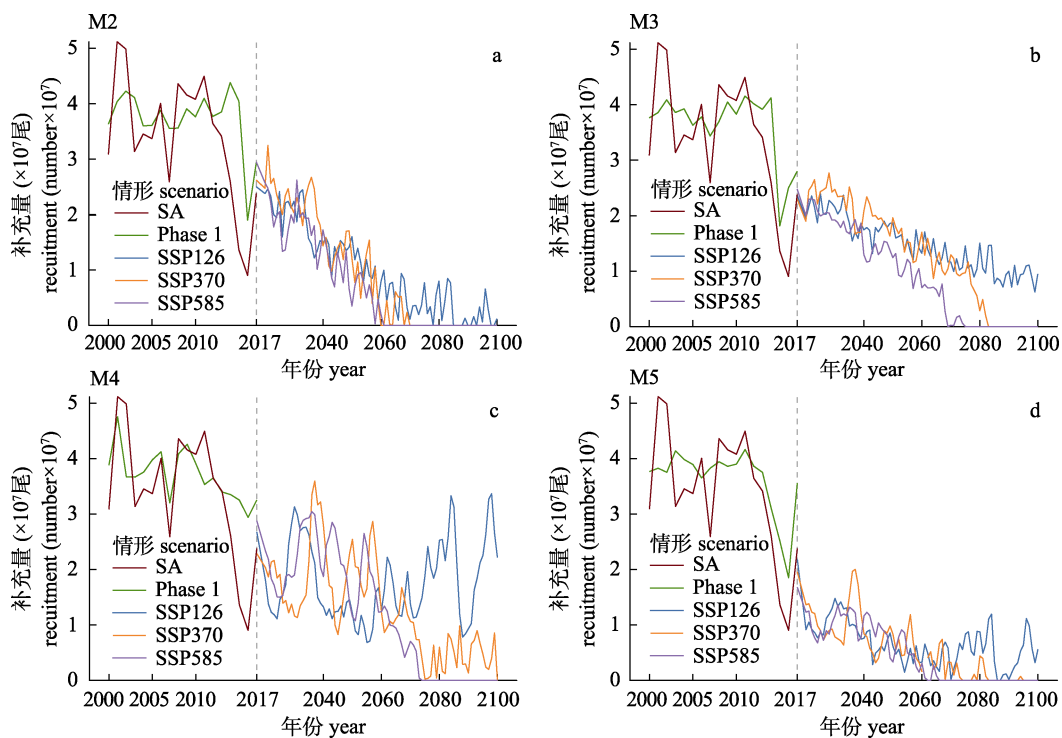


图 4 模型 M2 (a)、M3 (b)、M4 (c)、M5 (d) 对 2000—2100 年补充量预测结果

红线: 资源评估中提取的观测值; 绿线: 基于环境观测数据进行的第一阶段预测(2000—2017 年); 蓝线、黄线和紫线分别为 SSP126、SSP370、SSP585 情形下进行的第二阶段预测(2015—2100 年)。

Fig. 4 Recruitment predictions from 2000 to 2100 based on M2 (a), M3 (b), M4 (c), and M5 (d)
Red line (SA): observed recruitment extracted from the stock assessment; green line (phase 1): the first-stage predictions based on environmental observation data (2000–2017); blue, yellow, and purple lines correspond to the second-stage predictions (2015–2100) under the SSP126, SSP370, and SSP585 scenarios, respectively.

第一阶段中, M2 预测补充量与观测补充量趋势变化相似, 整体预测值相比观测值较低; M3 在 2000—2007 年未能预测出明显变化趋势, 2007 年之后预测趋势与观测值类似; M4 在 2000—2014 年预测趋势与观测值相似, 但 2014—2017 年相差较大; M5 整体预测值比观测值较低, 在 2008 年之后预测效果较好。第二阶段中, 补充量整体呈下降趋势, 其中二氧化碳高排放情形 SSP585 下降最快且波动剧烈, 二氧化碳低排放情形 SSP126 下降较缓。补充量在 2070 年左右在 SSP370 和 SSP585 情形下降为 0。M3 和 M4 相比 M2 和 M5 预测值较高, M5 相对预测值最低。

2.4 模型对比

LASSO 相较于 OLS 线性回归模型优化了变量选择过程, 以 M2 模型为例, 采用相同数据及假设, 对比 2 个模型发现, 基于 LASSO 回归方法的 M2 模型的均方根误差为 1.153, 比 OLS 回归低 (1.214); 解释方差为 0.459, 比 OLS 回归模型高 (0.431)(表 4)。

表 4 LASSO 方法与最小二乘法线性回归模型性能对比
Tab. 4 Comparison of LASSO and OLS linear regression model performance

模型 model	均方根误差 RMSE	解释方差 r-squared
M2	1.153	0.459
OLS 线性回归 OLS linear regression model	1.214	0.431

3 讨论

3.1 环境变量影响机制

本研究结果表明南太平洋长鳍金枪鱼补充量与气候变量之间关系存在较明显的空间异质性, 这与长鳍金枪鱼可进行体温调节, 在冷水中保持较高体温, 因此常分布于温带海域的生物学特征相吻合^[34]。以往研究表明, 长鳍金枪鱼有在夏季 SST 超过 24 °C 的热带和亚热带海域产卵, 后向温带海域迁徙觅食的特性^[23], 产卵区域环境虽决定了仔稚鱼的存活率, 但由于长鳍金枪鱼相对成熟较晚^[35], 生长至补充年龄更容易受温带区域环境影响, 支持了本研究中补充区域的环境数据

相比产卵区对补充量影响更大, 预测能力更强的观点。

SST 与长鳍金枪鱼产卵量和幼鱼存活率直接相关, 补充量在 SST 为 22~26 °C 时达到峰值, 超出此范围时呈指数下降^[36], 补充量和 SST 的对应关系与本研究结果一致, 研究表明 SST 是预测补充量的重要因素之一, 与补充量呈负相关, 且 SST 在各个区域重要性不同也反映了在不同区域所处的环境梯度不同。例如产卵区处于 SST 梯度的极端高值, SST 为最显著负相关变量; 而在补充区, SST 重要性降低。研究发现 MLD 也为预测补充量的显著因素, MLD 为海洋表层与深层水体之间的过渡区域, 其深度变化会影响海洋的温度结构、营养盐分布以及初级生产力, 浅层 MLD (<50 m) 通过增强营养盐垂向输送, 提升金枪鱼幼体的摄食成功率及存活率, 研究发现幼体胃含物丰度与 MLD 呈显著负相关, 当 MLD 超过 100 m 时, 真光层营养盐被稀释, 金枪鱼幼鱼每日摄食频率减少 30%, 导致生长率降低 18%~32%^[37], 此结果与本研究结果一致, MLD 与补充量成显著负相关关系, 但针对 MLD 与南太平洋长鳍金枪鱼关系的研究还十分有限。仔稚鱼生存主要依赖浮游动物, 因此 CHL 在产卵区影响较为显著。然而, 由于长鳍金枪鱼幼鱼食物主要包括凤尾鱼(*Engraulis mordax*)、沙丁鱼(*Sardinops sagax*)等鱼类^[38], CHL 在其他区域中重要性较低。SSS 是海水中营养盐指标, 其差异也会影响海水密度结构、海洋环流和路径, 因此 SSS 影响了鱼类多个生理过程, 如洄游、卵子受精、胚胎发育等^[39]。SSH 是反映海洋动力过程的重要指标, 引起底层营养盐和饵料鱼类聚集和分散^[40], SSH 的变化可能会影响鱼类产卵和繁殖活动的有效性从而影响鱼类种群的补充和丰度^[39]。研究结果显示 SSH 在 M2 和 M4 中对补充量有影响, 与补充量呈负相关关系, 但影响较低。

3.2 资源管理建议

基于气候-补充量关系进行渔业资源预测的研究仍存在较大不确定性, 例如气候预测模型偏差^[16]。相比 CMIP5, 本研究基于 CMIP6 气候模拟

情形对数据精度、海洋动力模型、大气-海洋作用等方面进行了改进,气候预测模型精度更高。预测结果发现,在较高二氧化碳排放 SSP585、SSP370 情形下,补充量可能受到严重影响,甚至面临种群崩溃的风险,SSP126 下虽风险较低,但补充量仍呈下降趋势。考虑到补充区域内除存在金枪鱼延绳钓渔船外,还存在一定规模的以金枪鱼为目标鱼种的曳绳钓渔业,该类渔船捕捞渔获的长鳍金枪鱼个体规格比延绳钓渔船更小,应当采用更为科学和精细的渔业管理措施。

不同气候情景下补充量的下降速度和程度存在差异,意味着渔业管理策略应更加灵活,以适应未来环境变化带来的不确定性。气候变化日益加剧情况下,渔业管理应及时调整总捕捞额度以保护种群及海洋生态系统可持续发展;预测模型也可识别高补充量窗口期,指导渔船集中作业于高产时期,进而提高渔获率并降低捕捞成本,同时还可评估不同管理策略下资源恢复潜力,优化中长期管理方案。

3.3 研究展望

渔业-气候模型中普遍存在过拟合现象,尤其当样本量较小时,这一问题更为明显^[41-42]。LASSO 方法通过对回归系数施加 L1 正则化约束,优化了模型变量选择,减少方差,从而降低了过拟合风险^[43]。与 OLS 多元线性回归对比发现,LASSO 解释方差更高,误差更低。在渔业管理中,即使模型最终性能指标差距很小,对最终结果也可产生较大偏差。其次,OLS 多元线性回归在数据训练集上可能过度拟合随机噪声,而 LASSO 方法通过压缩无关变量的系数,提高了模型对核心气候驱动因子的敏感性,在样本量较低的情况下,可避免因数据随机性导致的错误管理决策。LASSO 方法有效进行变量选择的优势,也可考虑应用于单位捕捞努力量(catch per unit effort, CPUE)标准化中,避免 GAM、GLM 等模型过拟合问题。本研究中尚未考虑补充量的密度制约、种间相互作用以及个体行为特征等生态系统内部变化,下一步研究可结合生态系统模型(如 Ecopath with Ecosim, EWE)预测生态系统在捕捞或环境变化下的动态响应,以提高模型精度。

参考文献:

- [1] Duffy-Anderson J T, Bailey K, Ciannelli L, et al. Phase transitions in marine fish recruitment processes[J]. *Ecological Complexity*, 2005, 2(3): 205-218.
- [2] He X, Field J C. Effects of recruitment variability and fishing history on estimation of stock-recruitment relationships: Two case studies from U.S. west coast fisheries[J]. *Fisheries Research*, 2019, 217: 21-34.
- [3] Johnston T A, Lester N P, Shuter B J. Recruitment[M]// Craig J F. *Freshwater Fisheries Ecology*. Oxford: Oxford University Press, 2015: 830-845.
- [4] Glória M B A, Daeschel M A, Craven C, et al. Histamine and other biogenic amines in albacore tuna[J]. *Journal of Aquatic Food Product Technology*, 1999, 8(4): 55-69.
- [5] 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.
- [6] Vidal T, Williams P, Ruaia T. Overview of tuna fisheries in the western and central Pacific Ocean, including economic conditions-2023[R]. Manila: Western and Central Pacific Fisheries Commission, 2024: 1-88.
- [7] Castillo-Jordan C C, Hampton J, Ducharme-Barth N, et al. Stock assessment of South Pacific albacore tuna[R]. New Caledonia: Western and Central Pacific Fisheries Commission, 2021: 1-108.
- [8] Cailliet M G, Andrews A H. Age-validated longevity of fishes: Its importance for sustainable fisheries[C]//Tsukamoto K, Kawamura T, Takeuchi T, et al. *Fisheries for Global Welfare and Environment*, 5th World Fisheries Congress. Tokyo: Terrapub, 2008: 103-120.
- [9] Chang Y J, Hsu J, Lai P-K, et al. Evaluation of the impacts of climate change on albacore distribution in the South Pacific Ocean by using ensemble forecast[J]. *Frontiers in Marine Science*, 2021, 8: 731950.
- [10] Singh A A, Sakuramoto K, Suzuki N, et al. Environmental conditions are important influences on the recruitment of North Pacific albacore tuna, *Thunnus alalunga*[J]. *Applied Ecology and Environmental Research*, 2017, 15(1): 299-319.
- [11] Novianto D, Susilo E. Role of sub surface temperature, salinity and chlorophyll to albacore tuna abundance in Indian Ocean[J]. *Indonesian Fisheries Research Journal*, 2016, 22(1): 17-26.
- [12] Stenseth N C, Ottersen G, Hurrell J W, et al. *Marine Ecosystems and Climate Variation: The North Atlantic-A Comparative Perspective*[M]. Oxford: Oxford University Press, 2004: 137-146.

- [13] Punt A E, A'Mar T, Bond N A, et al. Fisheries management under climate and environmental uncertainty: Control rules and performance simulation[J]. ICES Journal of Marine Science, 2014, 71(8): 2208-2220.
- [14] Bartolino V, Margonski P, Lindegren M, et al. Forecasting fish stock dynamics under climate change: Baltic herring (*Clupea harengus*) as a case study[J]. Fisheries Oceanography, 2014, 23(3): 258-269.
- [15] Zhang C, Li G, Chen X J. Impact of climate change on recruitment of *Trachurus murphyi* based on different regimes[J]. Haiyang Xuebao, 2021, 43(9): 48-58. [张畅, 李纲, 陈新军. 不同模式下气候变化对智利竹筴鱼补充量的影响[J]. 海洋学报, 2021, 43(9): 48-58.]
- [16] Wu S N, Chen X J. Relationship between the recruitment of the Pacific-cohort of chub mackerel (*Scomber japonicus*) and the influence factors on the spawning ground based on GLM and GAM[J]. Journal of Fisheries of China, 2020, 44(1): 61-70. [武胜男, 陈新军. 基于 GLM 和 GAM 的日本鲭太平洋群体补充量与产卵场影响因子关系分析[J]. 水产学报, 2020, 44(1): 61-70.]
- [17] Lin H, Wang J, Zhu J, et al. Evaluating the impacts of environmental and fishery variability on the distribution of bigeye tuna in the Pacific Ocean[J]. ICES Journal of Marine Science, 2023, 80(10): 2642-2656.
- [18] Xie M Y, Liu B, Chen X J. Prediction on fishing ground of *Ommastrephes bartramii* in Northwest Pacific based on deep learning[J]. Journal of Fisheries of China, 2024, 48(11): 61-72. [解明阳, 柳彬, 陈新军. 基于深度学习的西北太平洋柔鱼渔场预测[J]. 水产学报, 2024, 48(11): 61-72.]
- [19] Mondal S, Punt A E, Mendes D, et al. Teleconnection impacts of climatic variability on tuna and billfish fisheries of the South Atlantic and Indian Ocean: A study towards sustainable fisheries management[J]. Fish and Fisheries, 2025, 26(2): 240-256.
- [20] A'mar Z T, Punt A E, Dorn M W. The evaluation of two management strategies for the Gulf of Alaska walleye pollock fishery under climate change[J]. ICES Journal of Marine Science, 2009, 66(7): 1614-1632.
- [21] Hegyi G, Garamszegi L Z. Using information theory as a substitute for stepwise regression in ecology and behavior[J]. Behavioral Ecology and Sociobiology, 2011, 65(1): 69-76.
- [22] 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.
- [23] Kumar S, Attri S D, Singh K K. Comparison of Lasso and stepwise regression technique for wheat yield prediction[J]. Journal of Agrometeorology, 2019, 21(2): 188-192.
- [24] Czaja Jr R, Hennen D, Cerrato R, et al. Using LASSO regularization to project recruitment under CMIP6 climate scenarios in a coastal fishery with spatial oceanographic gradients[J]. Canadian Journal of Fisheries and Aquatic Sciences, 2023, 80(6): 1032-1046.
- [25] 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.
- [26] Huang B, Thorne P W, Banzon V F, et al. Extended reconstructed sea surface temperature version 5 (ERSSTv5): Upgrades, validations, and intercomparisons[J]. Journal of Climate, 2017, 30(20): 8179-8205.
- [27] NASA. MODIS-TERRA Level 3 Mapped chlorophyll data version R2018[DB/OL]. Greenbelt: NASA Ocean Biology DAAC, 2018. [2025-02-06]. <https://oceandata.sci.gsfc.nasa.gov/directdataaccess/>.
- [28] Climate Data Store. ORAS5 global ocean reanalysis monthly data from 1958 to present[EB/OL]. (2021-09-10) [2025-02-06]. <https://cds.climate.copernicus.eu/datasets/reanalysis-oras5?tab=overview>.
- [29] Potts J, Castillo-Jordan C, Day J, et al. Analysis of longline size frequency data for the 2024 South Pacific albacore and WCPO striped marlin assessments[R]. Manila: Western and Central Pacific Fisheries Commission, 2024:1-31.
- [30] Farley J H, Hoyle S D, Paige E J, et al. Maturity ogives for South Pacific albacore tuna (*Thunnus alalunga*) that account for spatial and seasonal variation in the distributions of mature and immature fish[J]. PLoS One, 2014, 9(1): e83017.
- [31] Farley J H, Williams A J, Hoyle S D, et al. Reproductive dynamics and potential annual fecundity of South Pacific albacore tuna (*Thunnus alalunga*)[J]. PLoS One, 2013, 8(4): e60577.
- [32] Hjort J. Fluctuations in the Great Fisheries of Northern Europe Viewed in the Light of Biological Research[M]. Copenhagen: International Council for the Exploration of the Sea, 1914: 1-228.
- [33] Wang Y, Zhang Y Y, Geng Z, et al. Impacts of phytoplankton availability on bigeye tuna (*Thunnus obesus*) recruitment in the Indian Ocean[J]. Frontiers in Marine Science, 2024, 11: 1392273.
- [34] Schaefer K M. Tuna: Physiology, Ecology, and Evolution[M]. San Diego: Academic Press, 2001: 225-270.
- [35] Gillett R. Tuna for tomorrow[R]. Mauritius: Indian Ocean Tuna Commission, 2013: 1-88.
- [36] Lehodey P, Bertignac M, Hampton J, et al. El Niño Southern Oscillation and tuna in the western Pacific[J]. Nature, 1997,

- 389(6652): 715-718.
- [37] Dueri S, Faugeras B, Maury O. Modelling the skipjack tuna dynamics in the Indian Ocean with APECOSM-E: Part 1. Model formulation[J]. Ecological Modelling, 2012, 245: 41-54.
- [38] Glaser S M. Interdecadal variability in predator-prey interactions of juvenile North Pacific albacore in the California Current System[J]. Marine Ecology Progress Series, 2010, 414: 209-221.
- [39] Mondal S, Ray A, Lee M A, et al. Projected changes in spawning ground distribution of mature albacore tuna in the Indian Ocean under various global climate change scenarios[J]. Journal of Marine Science and Engineering, 2023, 11(8): 1565.
- [40] Liu Y J, Zheng Q A, Li X F. Characteristics of global ocean abnormal mesoscale eddies derived from the fusion of sea surface height and temperature data by deep learning[J]. Geophysical Research Letters, 2021, 48(17): e2021GL094772.
- [41] Fox G A, Negrete-Yankelevich S, Sosa V J. Ecological Statistics: Contemporary Theory and Application[M]. Oxford: Oxford University Press, 2015: 44-57.
- [42] Maunder M N, Deriso R B. Estimation of recruitment in catch-at-age models[J]. Canadian Journal of Fisheries and Aquatic Sciences, 2003, 60(10): 1204-1216.
- [43] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction[M]. 2nd ed. New York: Springer, 2009: 219-259.

Prediction of South Pacific albacore *Thunnus alalunga* recruitment based on the LASSO regression method

WANG Yang^{1,2}, ZHU Jiangfeng^{1,2,3}, ZHANG Fan^{1,2,3}, GENG Zhe^{1,2*}

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

Abstract: Recruitment process serves as a critical biological foundation for sustainable resource maintenance. Understanding and accurately predicting the variability in recruitment has become a core challenge in fisheries management. Previous methods for predicting the recruitment of South Pacific albacore (*Thunnus alalunga*) have challenges in handling multicollinearity among environmental variables and identifying key drivers, often leading to model overfit and reduced predictive accuracy. To address these issues, this study applied the LASSO regression algorithm to optimize variable selection and improve prediction accuracy. Models were developed based on observational data from 2000 to 2017, and then coupled with CMIP6 multi-model climate projections, to predict recruitment trends from 2018 to 2100. The results indicated that LASSO effectively eliminated variable redundancy through shrinkage estimation, enhancing model prediction accuracy. The optimal model explained 45.9% of variance, with sea surface temperature (SST) and mixed layer depth (MLD) identified as critical predictors. Projections revealed that under high-emission scenarios (SSP5-8.5, SSP3-7.0), the population recruitment by the 2070s would approach the ecological threshold lower limit (near zero), significantly elevating collapse risks. In contrast, under the low-carbon pathway (SSP1-2.6), recruitment exhibited persistent decline trends. This study provided with an effective methodological framework for fisheries stock prediction and variable selection, while establishing a climate-recruitment coupled prediction model to provide quantitative decision-making support for formulating adaptive management strategies and mitigating population collapse risks.

Key words: recruitment prediction; LASSO method; CMIP6; albacore

Corresponding author: GENG Zhe. E-mail: zgeng@shou.edu.cn