#### DOI: 10.12264/JFSC2020-0503

# 基于 GAM 和权重分析的西北太平洋秋刀鱼渔情预报研究

刘瑜1,花传祥1,2

1. 上海海洋大学海洋科学学院, 上海 201306;

2. 国家远洋渔业工程技术研究中心, 上海 201306

摘要:为了提高秋刀鱼(Cololabis saira)渔情预报模型的时空分辨率,提升生产经济效益,本研究基于 2013—2016 年 7—11 月中国在西北太平洋公海的秋刀鱼生产数据及海洋环境数据,利用广义可加模型(generalized additive models, GAM)分别拟合单位捕捞努力量渔获量(catch per unit effort, CPUE)的适宜性指数(suitability index, SI)与各海洋环境变量之间的 SI 模型,结合提升回归树模型(boosting regression tree, BRT)进行权重分析,建立以月份为周期的秋刀鱼栖息地适宜性指数(habitat suitability index, HSI)模型。结果表明,(1) GAM 能较好地拟合适宜性指数与环境变量的关系,获得最优环境变量参数值;(2) 环境变量对 CPUE 影响权重的前 3 位分别为海表温度梯度、海表温度和混合层深度,其中,在秋季 9—11 月海表温度梯度的权重值均为最高;(3) HSI 模型的检验和评价总体准确率分别为 82.0%和 73.2%,秋季可达 87.7%和 77.9%,在盛渔期 10 月,预测准确率达 89.4%;(4) HSI 高值区与秋刀 鱼实际渔场在空间分布基本一致。研究表明该模型适用于秋刀鱼的渔情预报,并在每天的速报中具有明显优势。

关键词:秋刀鱼;权重分析;GAM;栖息地指数;渔情预报 中图分类号:S931 \_\_\_\_\_文献标志码:A \_\_\_\_\_文章编号:1005-8737-(2021)07-0888-08

秋刀鱼(Cololabis saira)为北太平洋中上层鱼 类,是我国重要的大洋性经济种类之一,也是北 太平洋渔业管理委员会优先管理鱼种之一<sup>[1]</sup>。由 于秋刀鱼的高度洄游性,在远洋渔业生产过程中, 渔船每天需变更生产海域,并投入大量成本寻找 渔场。但在目前国内外学者对秋刀鱼渔情预报的 研究中,预报的时间分辨率多为月或年<sup>[2-5]</sup>,针对 周或天的渔场速报研究非常少。作为一种 2 龄的 短生命周期鱼类,秋刀鱼的资源丰度和渔场分布 易受到不同海洋环境的影响<sup>[6-7]</sup>,因此,在预报模 型建立中,各个环境变量的权重分析十分重要。

提升回归树方法可以有效计算出环境变量对 模型建立的贡献大小<sup>[8]</sup>。广义可加模型是多元线 性回归的扩展,能够直接处理响应变量与多个解 释变量的非线性关系<sup>[9]</sup>。栖息地适宜性指数模型 是评价环境对生物影响的重要方法,在渔业领域 得到广泛的应用<sup>[10]</sup>,该模型的兼容性强,能够融 合 BRT 和 GAM 方法,既可以有效的选择环境变 量,也能够很好的分析渔场与海洋环境间的复杂 关系。本研究根据西北太平洋秋刀鱼生产数据和 海洋环境数据,结合 GAM 和环境变量的权重分 析,按月份构建秋刀鱼 HSI 模型,并分析模型在 秋刀鱼渔场预报中的可行性,以期为秋刀鱼生产 渔场速报提供参考。

# 1 材料与方法

#### 1.1 数据来源和预处理

渔业数据来自中国远洋渔业协会秋刀鱼技术 组,渔船主要作业海域为 38°~46°N, 150°~165°E, 数据时间为 2013—2016 年的 7—11 月,时间分辨

收稿日期: 2020-11-06; 修订日期: 2020-12-15.

基金项目: 国家重点研发计划项目(2020YFD0901203).

作者简介: 刘瑜(1985-), 女, 工程师, 从事渔业遥感研究. E-mail: liuy@shou.edu.cn

通信作者:花传祥,高级工程师,从事渔业海洋学研究.E-mail: cxhua@shou.edu.cn

率为天,数据项包括作业日期、经度、纬度、日 产量(t)和作业船数等信息,数据总量为 5620 条。 CPUE 定义为每艘船每天的渔获量,单位为 t/d, 作为渔业资源丰度的指标。采用渔场分布重心作 为资源量的空间分布和变化指标<sup>[11]</sup>,计算公式为:

$$X = \frac{\sum_{i=1}^{n} (C_{i} \times X_{i})}{\sum_{i=1}^{n} C_{i}}, \quad Y = \frac{\sum_{i=1}^{n} (C_{i} \times Y_{i})}{\sum_{i=1}^{n} C_{i}}$$
(1)

式中, *X*、*Y*分别为渔场作业重心的经度和纬度, *C<sub>i</sub>*为作业点*i*的 CPUE, *X<sub>i</sub>*和 *Y<sub>i</sub>*分别为作业点*i*的 纬度和经度, *n*为作业总次数。

环境数据首先选用海表面温度(sea surface temperature, SST)及其变化海表温度梯度(sea surface temperature gradient, SSTG),其通常作为判 断渔场的关键指标<sup>[6]</sup>,来源于美国国家海洋和大 气管理局网站(http://www.noaa.gov),其空间分辨 率为  $0.01^{\circ} \times 0.01^{\circ}$ 。SSTG 利用梯度幅值公式求得<sup>[12]</sup>, 假设当前 SST 值为 SST<sub>*i*,*j*</sub>,其相邻的 4 个网格点的 SST 分别为 SST<sub>*i*+1,*j*</sub>、SST<sub>*i*-1,*j*</sub>、SST<sub>*i*,*j*+1</sub>、SST<sub>*i*,*j*-1</sub>, 则

 $SSTG_{i,j} =$ 

$$\left(\frac{\text{SST}_{i+1,j} - \text{SST}_{i-1,j}}{\Delta x}\right)^2 + \left(\frac{\text{SST}_{i,j+1} - \text{SST}_{i,j-1}}{\Delta y}\right)^2 (2)$$

式中,*i、j*为正整数,分别表示网格数据行号和列 号, Δ*x* 表示第 *j*-1 列与第 *j*+1 列之间经度方向的 距离, Δ*y* 表示第 *i*-1 行与第 *i*+1 行之间纬度方向 的距离, 单位均为 km。

其次是叶绿素 a 浓度(chlorophyll-a concentration, Chl-a), 其可以反映渔场的初级生产力条 件,来自哥白尼海洋环境监测中心(http://marine. copernicus.eu), 空间分辨率为 0.25°×0.25°。海面 高度异常(sea level anomaly, SLA)和涡动能(eddy kinetic energy, EKE) 则可以作为动力环境和海流 活跃程度的指示因子<sup>[13]</sup>,来自卫星高度计资料, 取自法国空间局 AVISO 网站(http://www.aviso. oceanobs.com), 其中包括海面高度数据和地转流 数据,空间分辨率 0.25°×0.25°, 涡动能的计算公 式为<sup>[13]</sup>:

EKE = 
$$\frac{1}{2}(u'^2 + v'^2)$$
 (3)

式中, EKE 为涡动能, 单位为 cm<sup>2</sup>/s<sup>2</sup>;  $u' = u - \overline{u}$ ,  $v' = v - \overline{v}$ ,  $\overline{u}$  和 $\overline{v}$  分别是纬向速度 u 和经向速度 v的年平均值。最后, 混合层深度(mixed layer depth, MLD)可以作为海水垂直分层的指示因子<sup>[6]</sup>, 来自 全球海洋 HYCOM 模式数据(https://www.hycom. org), 空间分辨率 1°/12×1°/12。本研究采用的环境 数据时间分辨率均为天。

根据渔船作业的日期和经纬度信息,查找相 应日期海洋环境数据网格上与该作业位置距离最 近的经纬度所对应的数据值,匹配为该作业位置 的环境数据经纬度和环境数据值。

# 1.2 研究方法

**1.2.1 权重分析** BRT方法是在运算过程中随机 抽取一定量的数据分析自变量对因变量的影响程 度,而剩余数据用来对拟合结果进行检验,对生 成的多重回归树取平均值输出,从而得出自变量 对因变量的影响大小<sup>[8]</sup>。本文 BRT 模型的数据计 算采用了 R 软件中的梯度提升函数(gradient boosting machine, GBM)<sup>[14]</sup>,其中抽样率(train. fraction)设置为0.8,重复循环计算1000次,获取环境 变量的权重分布,并取平均值计算标准差进行 检验。

**1.2.2** SI 模型 本研究利用 GAM 模型的样条平 滑函数建立各个环境变量与 SI 的关系模型, 表达 形式如下:

 $\ln(SI) = \alpha + f(x_i) + \varepsilon \tag{4}$ 

式中, SI 为变量的适宜性指数;  $\alpha$  为模型的截距; f 为非参数平滑函数;  $x_i$ 为第 i 个解释变量;  $\varepsilon$ 为 残差,  $\varepsilon = \sigma^2 \pm E(\varepsilon) = 0$ 。模型采用的是样条平滑 法,误差分布估计为高斯分布<sup>[15]</sup>。通常研究以 CPUE 建立 SI, 假定最高 CPUE 为资源丰度最大 的海域。但秋刀鱼的生产数据每天的差距量级可 达 100,直接使用 CPUE 会降低 SI 指数精度。因 此本文对每个环境变量进行数据分组,计算每个 环境变量在各组距范围内的 CPUE 累积频率,建 立每个组距下的 SI, 计算公式为:

$$SI_m = \frac{CPUE_m}{CPUE_{max}}$$
(5)

%

式中, SI<sub>m</sub> 为环境变量在每个组距范围内的适宜性 指数, CPUE<sub>m</sub> 为每个组距范围下的 CPUE 累积频 率, CPUE<sub>max</sub> 为每个组距范围下的最大 CPUE 累积 频率。根据每个组距下的 SI 和环境数据进行样条 平滑函数拟合,依据拟合优度 *R*<sup>2</sup>,确定各个环境 变量的最适组距,并进行显著性 *F* 检验,当 *P*<0.01 时,则通过显著性检验,并依次分别求出 各变量的 SI 指数。

**1.2.3 HSI 模型** 结合权重分析和 SI 指数,本文利用赋予权重的算数平均算法建立 HSI 模型,算数平均算法具有不受 SI 极值影响的特点,是较为常见的有效方法,计算公式如下:

$$HSI = \sum_{i=1}^{n} W_i SI_{mi}$$
(6)

式中, *i* 为第 *i* 个环境变量, *n* 为环境变量总数, *W<sub>i</sub>* 为第 *i* 个环境变量的权重, SI<sub>mi</sub> 为第 *i* 个环境变量的 SI 值。HSI 取值范围一般为 0~1, 0 表示不适宜栖息地, 1 表示最适宜栖息地<sup>[10]</sup>。本文利用 2013 —2015 年 7—11 月的数据分月建立 HSI 模型, 统计以 0.1 为间距的 HSI 值中产量占当月总产量的比重和平均 CPUE, 对模型进行检验。然后将 2016 年 7—11 月的环境数据分别输入模型, 得到 HSI 预报值, 并绘制 2016年生产作业点与 HSI 指数的空间分布图, 以对渔场预报效果进行评价。

#### 2 结果与分析

#### 2.1 环境变量选择

根据 BRT 方法分别计算各月环境变量 SST、 SSTG、Chl-a、EKE、SLA 和 MLD 对 CPUE 的影 响,结果显示,各个环境变量的权重标准差均小 于 0.05,故可以使用平均值代表权重大小(表 1), 不同环境变量对 CPUE 的拟合权重贡献率分布在 10.9%~20.6%之间,7月最高和最低权重相差最大 为 9.4%。不同的月份权重分布各不同,8—10月排 在前 3 位的变量分别为 SSTG、SST、MLD,7月 为 SSTG、MLD、EKE,11月为 SSTG、Chl-a、SST, 其中 SSTG 在秋季 9—11月的权重值均为最高, 而第 4 位之后各月权重排名均不相同。考虑到环 境变量之间的相互作用会对模型产生影响,因此 本文选择各月权重排在前 4 位的变量建立模型, 以减少这种相互作用,并重新计算权重贡献率,结果见表2。

表 1 筛选变量的权重 Tab. 1 Weight of selected parameter

月份 month	W <sub>sst</sub>	W <sub>sstg</sub>	$W_{ m chla}$	W <sub>sla</sub>	W <sub>eke</sub>	$W_{ m mld}$
7	14.25	19.20	16.35	13.45	16.46	20.30
8	16.99	19.76	10.88	15.95	16.13	20.30
9	17.88	19.30	15.87	14.12	13.62	19.21
10	18.36	20.56	14.16	15.73	14.99	16.19
11	17.18	17.87	17.61	15.57	14.67	17.10

表 2 各变量的 SI<sub>m</sub> 拟合结果 Tab. 2 Result of fitted SI<sub>m</sub>

月份 month	环境变量 parameter	F	$R^2$	组距步长 step interval	权重/% weight
7	SSTG	23.7	0.86	0.007 °C/km	27.06
July	EKE	35.11	0.93	$11 \text{ cm}^2/\text{s}^2$	22.91
	Chla	10.21	0.69	$0.08 \text{ mg/m}^3$	22.71
	MLD	22.22	0.68	2 m	27.32
8	SST	117	0.98	0.6 °C	23.98
August	SSTG	42.22	0.81	0.002 °C/km	26.16
	EKE	104.5	0.97	$11 \text{ cm}^2/\text{s}^2$	22.36
	MLD	36.72	0.92	3.1 m	27.50
9	SST	278.4	0.98	0.8 °C	24.88
September	SSTG	39.62	0.94	0.015 °C/km	26.91
	Chla	17.31	0.76	$0.07 \ mg/m^3$	21.36
	MLD	19.17	0.93	4 m	26.85
10	SST	100.6	0.98	1 °C	25.21
October	SSTG	27.4	0.93	0.018 °C/km	27.62
	SLA	67.88	0.96	2.5 cm	24.41
	MLD	52.86	0.97	5.9 m	22.76
11	SST	56.83	0.98	1 °C	25.16
November	SSTG	181.9	0.98	0.022 °C/km	25.07
	Chla	30.65	0.77	$0.15 \ mg/m^3$	25.06
	MLD	17.89	0.90	5.8 m	24.72

#### 2.2 模型建立

结合权重分析结果,依据样条平滑函数拟合 优度 R<sup>2</sup>、自由度、F 值以及显著性检验,循环计 算得出 7—11 月各个环境变量的组距步长,结果 见表 2。结果显示,环境变量均在 P<0.01 水平下 影响显著,通过显著性 F 检验。环境变量拟合方 程的拟合度 R<sup>2</sup>对于 SST、SSTG、EKE、SLA 均 达到 0.8 以上,其中 Chl-a 在 7 月、9 月和 11 月分 别为 0.69、0.76、0.77,7 月的 MLD 为 0.68,较高 的 *R*<sup>2</sup> 说明 GAM 能够较好地拟合 SI 与各个环境变 量的关系。

结合 7—11 月环境变量的 SI 拟合指数和各自 对应的权重大小,运用算数平均值方法建立 HSI 模型,分别如下:

$$\begin{split} HSI_7 &= SI_{mld} \cdot 27.32\% + SI_{sstg} \cdot 27.06\% + \\ SI_{eke} \cdot 22.91\% + SI_{chla} \cdot 22.71\% \\ HSI_8 &= SI_{mld} \cdot 27.5\% + SI_{sstg} \cdot 26.16\% + \\ SI_{sst} \cdot 23.98\% + SI_{eke} \cdot 22.36\% \\ HSI_9 &= SI_{sstg} \cdot 26.91\% + SI_{mld} \cdot 26.85\% + \\ SI_{sst} \cdot 24.88\% + SI_{chla} \cdot 21.36\% \\ HSI_{10} &= SI_{sstg} \cdot 27.62\% + SI_{sst} \cdot 25.21\% + \end{split}$$

$$SI_{sla}\cdot 24.41\% + SI_{mld}\cdot 22.76\%$$

$$\begin{split} HSI_{11} = &SI_{sst} \cdot 25.16\% + SI_{sstg} \cdot 25.07\% + \\ &SI_{chla} \cdot 25.06\% + SI_{mld} \cdot 24.72\% \end{split}$$

#### 2.3 模型检验与评价

通过对不同 HSI 组距下,秋刀鱼的产量比重 和平均 CPUE 的计算,图 1 结果显示,2013—2015 年7—11月的 HSI 值在 0.5~0.9 范围内的产量比重 和分别为 68.5%、75.6%、90.8%、86.6%、88.7%, 平均为 82.0%,其中秋季 9—11 月平均为 87.7%; HSI 值在 0.5、0.6、0.7、0.7、0.7 时产量比重最 高,分别为 25.5%、28.8%、27.7%、23.5%、25.5%, 各月最高 CPUE 对应的 HSI 值分别为 0.8、0.6、 0.6、0.7、0.7。由 2016 年 7—11 月的生产数据对 模型进行评价,计算得到 HSI 值在 0.5~0.9 范围内 的产量比重和,分别为 99.4%、32.9%、67.3%、 89.4%、76.9%,平均为 73.2%,其中秋季 9—11



Fig. 1 Mean CPUE and percentages of yield under different HSI values

月平均为 77.9%; HSI 值在 0.6、0.4、0.5、0.8、 0.5 时的最高,产量比重分别为 27.5%、51.5%、 35.0%、28.2%、34.1%,各月最高 CPUE 对应的 HSI 值分别为 0.6、0.4、0.5、0.8、0.4。总体来看, 7—9 月的预报结果与模型检验相差较大,8 月的 预报效果最差。秋季 9—11 月比夏季 7—8 月的 HSI 模型效果好,秋季为秋刀鱼的盛渔期,2013— 2016 年 10 月的平均 CPUE 达 23.7 t/d,而 8 月的 平均 CPUE 为 7.3 t/d。在秋季,以 10 月效果最佳, 其预报 HSI 超过 0.5 以上时产量比重达 89.4%, 且 HSI 值在 0.2~0.8 范围时, 随着 HSI 指数的增加, CPUE 呈逐渐上升趋势。本文以 HSI 值大于 0.5 为适宜秋刀鱼渔场形成海域, HSI 值大于 0.7 为高值 区,由以上结果可见,秋刀鱼高产渔区基本集中在 HSI 指数大于 0.5 的范围内,模型效果良好。

本研究利用各月的预报模型分别绘制 2016 年7—11月各月 20—22 日的 HSI 分布图,并叠加 匹配各月每天的生产作业位置,结果如图 2 所示。



图 2 2016 年 7—11 月各月 20—22 日秋刀鱼 HSI 和作业位置(红色圆圈)分布 Fig. 2 The distribution of HSI and CPUE in the fishing ground (red circle) of Pacific saury from July to November

7月HSI高值区主要分布在45°N以北和160°E以 东海域,尽管HSI 值平均在 0.6 左右, 但实际生产 点基本落在 HSI 高值范围, 故 7 月预测的产量比 重为最高月份。8月 HSI 高值区向西南方向移动, 同时高值范围的面积逐渐减少, 而渔船则向南移 动越过了 HSI 高值区, 作业位置完全分布于 HSI 小于0.5区域内,故8月预测的产量比重为最低月 份。9月 HSI 高值区继续向南移动, 高值面积集 中且与渔船分布均呈条带状分布, 作业位置集中 落在高值区域,因此,尽管9月HSI值平均为0.5, 但高值范围内的产量比重高于8月。10月为秋刀 鱼盛渔期, 整个海域 HSI 高值海域面积范围广, 作业位置集中在 42°N 和 153°E 附近的海域, 且基 本分布于 HSI 值大于 0.8 的范围内, 因此预测结 果最佳。11 月 HSI 高值区主要分布于 42°N 和 155°E 以东的海域, 而作业位置仍位于 155°E 以 西的高值区域, 故产量比重较 10 月低。总体上看, HSI 高值区域和实际渔场的位置较为接近, 且二 者的空间分布移动方向基本一致。

#### 3 讨论

### 3.1 环境变量对渔场分布的影响

作为短周期高度洄游鱼类,秋刀鱼对温度具 有高度的敏感性,且其洄游路线很长。一般在夏 末秋初,秋刀鱼从亲潮水域南下向黑潮水域移动, 幼鱼逐渐成熟,秋季移动至黑潮附近,黑潮暖水 与亲潮冷水交汇,海洋动力过程活跃,往往形成 温度锋面现象,引起海水垂直混合加强,营养盐 富集并伴有浮游植物大量繁殖,形成高生产力区, 从而形成秋刀鱼盛渔期渔场[16]。在整个洄游过程 中,秋刀鱼对 SST 的响应更多呈现出季节性变化 的特点。而在盛渔期,鱼群集聚,捕捞作业范围缩 小,以往研究也发现,秋刀鱼作业海域的温度锋 面越活跃, 作业位置距离锋面越近, 单位渔获量 越高<sup>[17-18]</sup>,因此,在秋季 SSTG 对渔场分布的影 响更加突出。本研究以时间分辨率为月份的统计 分析中,SSTG和SST的权重值排序与秋刀鱼的洄 游集群特点基本一致。

其次,本研究发现,MLD 的权重贡献率基本 在前3位,这与秋刀鱼对 SSTG 的响应特征情况 相似。由于秋刀鱼的垂直洄游特性,即白天分布 水深为 30~40 m 水层,夜间为 0~5 m 的近表层<sup>[19]</sup>, 而在温度锋面处由于海水的垂直混合从而 MLD 加大,这一变化对应着秋刀鱼可适应的上层水体 向下延伸,从而扩展秋刀鱼的上下活动范围。因 此,MLD 的变化影响鱼群的栖息水层分布,进而 影响了鱼群集聚和渔场分布。第三,对于 Chl-a、 SLA 和 EKE 3 个环境变量,主要体现在叶绿素、 海流、涡旋等对鱼群的影响,在不同的月份由于 鱼群成熟期不同,因而影响各不相同;在短期尤 其盛渔期期间,可能与这 3 个变量在渔场区域的 变化幅度相对于 SST、SSTG、MLD 较小有关,因 而对渔场分布的影响权重贡献率略低。

## 3.2 HSI 建模及对渔场分布的影响

本研究利用 GAM 方法的样条平滑函数结合 环境变量的权重分析结果,分月建立了 HSI 模型, 并分别对模型进行了检验和评价,总体精度分别 为82.0%和73.2%,其中秋季可达87.7%和77.9%。 但7月和8月模型检验和评价结果相差较大,导 致这个现象的原因可能是由于, 在夏季, 秋刀鱼 处于成长期, 在洄游过程不断改变方向寻找适宜 温度等适宜环境,中心渔场不确定,这导致渔船 需要花费大量时间寻找渔场, 当遇到捕捞产量减 少时,出于经济效益考虑,渔船往往随之迅速转 移渔场, 使得实际作业时间较短, 渔场位置也在 不断变化中。由于本文采用生产数据,作业位置 一般会聚集在较小的空间区域,可能导致局部环 境变量的权重被放大,从而降低了非传统作业区 域的预报精度。此外,不同年份下的大尺度气候 模态不同, 气候模态通常直接影响温度的位置变 化,会导致渔场在这一年整体偏南或者偏北<sup>[5]</sup>, 因此,本研究的模型在夏季预报中不稳定,对于 7月和8月的预测并不适用。但模型在秋季的预 测效果良好, 尤其是 10 月, 预测精度达 89.4%, 这是因为在盛渔期,秋刀鱼在温度季节性变化以 及其洄游特性的影响下, 向南迁移集中至黑潮区 域,这一区域环境变量(SSTG、MLD等)的改变有 利于秋刀鱼的生存,本研究中,权重分析的结果 亦是反映出了这一变化结果,因此,该模型在短 期速报中具有优势,适用于秋季集中作业期间,

能更快地确定渔场方向和范围。

相较于以往神经网络、最大熵以及基于灰色 系统等西北太平洋秋刀鱼预测模型来说<sup>[6,20-21]</sup>, 本研究通过HSI模型的算法很好的融合了BRT和 GAM 方法,体现出了不同环境变量的重要性以 及减少了各个环境变量之间的相互影响,且该方 法灵活易操作,具有较强的实用性。但本研究的 栖息地预报准确度总体而言仍不够精确,模型在 盛渔期每天的预测中具有优势,在其他时期有一 定的局限性,需要进一步将大尺度环境变量如厄 尔尼诺、南方涛动等加入分析,以及根据秋刀鱼 的洄游和集群规律,进一步开展不同时间的局 部海洋动力环境特征研究,以提高西北太平洋 秋刀鱼渔场速报精度,提高远洋渔业生产效率 和经济效益。

#### 参考文献:

- Zavolokin A. Priority species[EB/OL]. [2018-06-09]. https:// www.npfc.int/priority-species.
- [2] Meng L W. Study on fishery forecast research of *Cololabis saira* in north Pacific Ocean based on habitat model[D].
   Shanghai: Shanghai Ocean University, 2017. [孟令文. 基于 栖息地指数的北太公海秋刀鱼渔情预报的应用研究[D].
   上海: 上海海洋大学, 2017.]
- [3] Hua C X, Zhu Q C, Shi Y C, et al. Comparative analysis of CPUE standardization of Chinese Pacific saury (*Cololabis saira*) fishery based on GLM and GAM[J]. Acta Oceanologica Sinica, 2019, 38(10): 100-110.
- [4] Zhu W T, Chen X J, Wang J T, et al. Predicting the abundance of Pacific saury based on Grey System in the Northwest Pacific[J]. Journal of Guangdong Ocean University, 2018, 38(6): 13-17. [朱文涛,陈新军,汪金涛,等. 基于灰色系统的西北太平洋秋刀鱼资源丰度预测[J]. 广东海洋大学学报, 2018, 38(6): 13-17.]
- [5] Chang Y J, Lan K W, Walsh W A, et al. Modelling the impacts of environmental variation on habitat suitability for Pacific saury in the Northwestern Pacific Ocean[J]. Fisheries Oceanography, 2019, 28(3): 291-304.
- [6] Hua C X, Zhu Q C, Xu W, et al. Review of the life history, resources and fishing grounds of the Pacific saury in the North Pacific Ocean[J]. Journal of Fishery Sciences of China, 2019, 26(4): 811-821. [花传祥,朱清澄,许巍,等. 北太平 洋秋刀鱼生活史和资源渔场研究进展[J]. 中国水产科学, 2019, 26(4): 811-821.]

- [7] Tian Y J, Akamine T, Suda M. Modeling the influence of oceanic-climatic changes on the dynamics of Pacific saury in the northwestern Pacific using a life cycle model[J]. Fisheries Oceanography, 2004, 13: 125-137.
- [8] De'ath G. Boosted trees for ecological modeling and prediction[J]. Ecology, 2007, 88(1): 243-251.
- [9] Berg D. Bankruptcy prediction by generalized additive models[J]. Applied Stochastic Models in Business and Industry, 2007, 23(2): 129-143.
- [10] Gong C X, Chen X J, Gao F, et al. Review on habitat suitability index in fishery science[J]. Journal of Shanghai Ocean University, 2011, 20(2): 260-269. [龚彩霞,陈新军, 高峰,等. 栖息地适宜性指数在渔业科学中的应用进展[J]. 上海海洋大学学报, 2011, 20(2): 260-269.]
- [11] 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.
- [12] Liu Y, Zheng Q A, Li X F. Relationship between neon flying squid Ommastrephes bartramii fishery distribution patterns and eddy kinetic energy in Northwest Pacific Ocean[J]. Haiyang Xuebao, 2020, 42(2): 44-51. [刘瑜,郑全安,李晓峰. 西北太平洋柔鱼渔场分布与涡动能变化的相关关系[J]. 海洋学报, 2020, 42(2): 44-51.]
- [13] Pi Q L, Hu J Y. Analysis of sea surface temperature fronts in the Taiwan Strait and its adjacent area using an advanced edge detection method[J]. Science China: Earth Sciences, 2010, 53(7): 1008-1016.
- [14] Elith J, Leathwick J R, Hastie T. A working guide to boosted regression trees[J]. Journal of Animal Ecology, 2008, 77(4): 802-813.
- [15] Howell E A, Kobayashi D R. El Niño effects in the Palmyra Atoll region: Oceanographic changes and bigeye tuna (*Thunnus obesus*) catch rate variability[J]. Fisheries Oceanography, 2006, 15(6): 477-489.
- [16] Kurita Y. Energetics of reproduction and spawning migration for Pacific saury (*Cololabis saira*)[J]. Fish Physiology and Biochemistry, 2003, 28(1-4): 271-272.
- [17] Tseng C T, Sun C L, Belkin I M, et al. Sea surface temperature fronts affect distribution of Pacific saury (*Cololabis saira*) in the Northwestern Pacific Ocean[J]. Deep Sea Research Part II: Topical Studies in Oceanography, 2014, 107: 15-21.
- [18] Liu Y, Zheng Q A, Li X F. Relationship between *Cololabis saira* fishery distribution patterns and sea surface temperature front in the Northwestern Pacific Ocean[J]. Journal of Fisheries of China, 2018, 42(12): 1916-1926. [刘瑜, 郑全安, 李晓峰. 西北太平洋公海秋刀鱼渔场分布与海表温度锋

的相关关系[J]. 水产学报, 2018, 42(12): 1916-1926.]

- [19] Hatanaka M, Watanabe T. Studies on the reproduction of the saury, *Cololabis saira* (Brevoort), of the Pacific coast of Japan[J]. Tohoku Journal of Agricultural Research, 1953, 3(2): 293-302.
- [20] Syah A F, Saitoh S I, Alabia I D, et al. Detection of potential fishing zone for Pacific saury (*Cololabis saira*) using gener-

alized additive model and remotely sensed data[J]. IOP Conference Series: Earth and Environmental Science, 2017, 54: 012074.

[21] Hua C X, Li F, Zhu Q C, et al. Habitat suitability of Pacific saury (*Cololabis saira*) based on a yield-density model and weighted analysis[J]. Fisheries Research, 2020, 221: 105408.

# Forecasting Pacific saury (*Cololabis saira*) fisheries based on GAM and weighted analysis in the northwest Pacific

LIU Yu<sup>1</sup>, HUA Chuanxiang<sup>1, 2</sup>

1. College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China;

2. National Engineering Research Centre for Oceanic Fisheries, Shanghai 201306, China

**Abstract:** To improve the spatial and temporal resolution of fishery forecast models and the resource utilization and economic benefits of Pacific saury (*Cololabis saira*), a generalized additive model (GAM) was used to fit the suitability index between the catch per unit effort (CPUE) and marine environmental variables, based on Chinese saury fishery and environmental data from the high seas of the northwest Pacific Ocean during July and November from 2013 to 2016. Weighted analysis was also conducted using boosted regression tree models to develop monthly habitat suitability index (HSI) models. The results indicated that the GAM can be reliably used to fit relationships between the suitability index and environmental variables and can obtain optimal environmental variable values. Weighted analysis showed that the three important environmental variables affecting CPUE were sea surface temperature gradient, sea surface temperature, and mixed layer depth. The weight of the sea surface temperature gradient was the highest during September to November (autumn). The overall accuracy of the HSI model test and evaluation stages were 82.0% and 73.2% respectively, reaching 87.7% and 77.9% in autumn, respectively. Furthermore, forecast accuracy was 89.4% in October during the main fishing season. The high-HSI areas were consistent with the fishing grounds of Pacific saury. Thus, the results show that the HSI model is suitable for forecasting the saury fishery and has a significant advantage in daily forecasting.

Key words: *Cololabis saira*; weighted analysis; GAM; habitat suitability index; fishery forecast Corresponding author: HUA Chuanxiang. E-mail: cxhua@shou.edu.cn