

文章编号: 1000-0615(2018)06-0889-13

DOI: 10.11964/jfc.20170610878

## 海州湾双斑螯栖息分布特征与环境因子的关系

栾静<sup>1</sup>, 张崇良<sup>1\*</sup>, 徐宾铎<sup>1</sup>, 薛莹<sup>1</sup>, 任一平<sup>1,2</sup>

(1. 中国海洋大学水产学院, 山东青岛 266003;

2. 青岛海洋科学与技术国家实验室, 海洋渔业科学与食物产出过程功能实验室, 山东青岛 266237)

**摘要:** 为了解双斑螯栖息分布规律, 实验根据2011—2016年多个季度航次在海州湾进行的渔业资源和环境调查数据, 采用广义线性模型(GLM)、广义可加模型(GAM)以及随机森林3种物种分布模型(SDMs)方法, 结合AIC(akaike information criterion)准则、累积偏差解释率和交叉检验等评判指标筛选和构建了双斑螯栖息分布模型, 并分析了环境因子对双斑螯分布的影响。结果显示, 3种模型在解释因子与响应变量间的关系上基本一致; 其中GAM在模型拟合上具有优势, 而随机森林的预测性能明显高于传统的GLM和GAM。双斑螯相对渔获量在年份和月份间的变异性最为显著, 两个因子的解释率分别在18%和3.8%以上。水深和表层盐度对双斑螯资源分布的影响较大, 均与双斑螯相对丰度呈正相关关系; 双斑螯分布总体呈现冬季相对较高, 夏季东北部海域高、西南部低的特点, 与海州湾水深分布特点基本一致。本研究还根据FVCOM(finite-volume coasta ocean model)模拟环境数据, 利用随机森林分布模型估计了双斑螯在海州湾海域2011年各个季节的空间分布, 为渔业资源的开发和保护提供依据。

**关键词:** 双斑螯; 广义线性模型(GLM); 广义可加模型(GAM); 随机森林; 模型比较; 交叉检验

中图分类号: S 931.1

文献标志码: A

双斑螯(*Charybdis bimaculata*)属于梭子蟹科(Portunidae)、螯属(*Charybdis*), 在我国的东海、黄海、渤海海域均有分布<sup>[1-2]</sup>, 在东、黄海区蟹类群落结构中占据优势种地位<sup>[1, 3]</sup>。双斑螯既是黄海近岸底栖生物食性功能群中的主要种类<sup>[4-5]</sup>, 也是许多高营养级鱼类的饵料生物<sup>[6]</sup>, 因此在黄海食物网以及底栖生物群落中具有重要地位。国内外关于双斑螯的研究已有报道, 主要涉及生物学特征<sup>[3, 7]</sup>、种群动力学<sup>[8]</sup>、空间分布与环境因子关系<sup>[2]</sup>以及分子系统地理学研究<sup>[1]</sup>等方面。近几年来随着人类活动和气候变化, 海洋栖息地发生了变化<sup>[9-10]</sup>, 导致许多物种资源量发生衰退和空间分布的变化, 进而可能对食物网以及生态系统结构和功能产生影响。因此, 进一步研究双斑螯资源空间分布与因子的关系,

预测其资源分布和变化趋势, 对渔业管理具有重要意义<sup>[11]</sup>。

物种的时空分布与环境因子之间的关系一直是生态学研究热点问题。物种分布模型(SDMs)是一类主要的研究方法, 它是基于生态位理论, 对物种分布数据以及环境或空间特征进行关联, 分析得到物种分布与环境变量或空间特征的关系, 用来估计物种真实或潜在分布区<sup>[12]</sup>。在渔业领域应用较为广泛的SDMs是广义线性模型(GLM)、广义可加模型(GAM)和栖息地指数模型; 除此之外, 一些机器学习方法在筛选重要因子和渔场预测上显示一定的优势<sup>[13-14]</sup>。SDMs研究主要涉及资源分布与环境因子关系<sup>[9, 15]</sup>、渔场预报<sup>[13]</sup>、CPUE标准化<sup>[14, 16-17]</sup>等。在海洋渔业研究中, 渔业数据呈现出一定的模糊性、复杂

收稿日期: 2017-06-16 修回日期: 2017-10-11

资助项目: 中央高校基本科研业务费专项(201612002, 201712015)

通信作者: 张崇良, E-mail: zhangclg@ouc.edu.cn

性和不确定性<sup>[13]</sup>，不同模型对数据的拟合和预测程度不同，需要应用合适的模型来解释变量间的关系和进行预测。相关研究大多以AIC (akaike information criterion)等准则评判模型的拟合程度并进行模型筛选，而对评估模型的预测性能研究较少。另外，有关物种分布研究多集中于经济型鱼类，而基于生态系统中贡献率高的饵料类研究较少。

本实验采用传统的GLM、GAM模型和基于机器学习的随机森林模型(Random Forest)，分析了双斑螭的资源分布与空间环境的关系，探讨影响双斑螭资源分布的主要因子，为渔业管理和生态系统模型的生态假设提供依据；评价了模型的拟合效果和预测性能，为预测资源分布提供技术参考，研究结果对渔业管理策略制定和海洋保护区选划有指导意义。

## 1 材料与方法

### 1.1 数据来源

双斑螭资源量数据来自于2011—2016年海州湾春、秋季以及2011年3月、8月、12月和2015年8月的渔业资源底拖网调查，调查海域为

119°20'E~121°10'E、34°20'N~35°40'N，该海域采用分层随机取样的方法进行调查(图1)。每个航次随机选取一定的站位进行调查(每10'×10'的方格内设置1个站位)。其中，2011年各航次分别选取24个调查站位；其他年份各航次分别选取18个调查站位。调查用船为220 kW的单拖渔船，拖速为2~3 kn。调查网具的网口宽度25 m，囊网网目17 mm。渔获样品带回实验室处理和统计。并进行拖网时间(1 h)以及拖速(2 kn)的标准化处理，得到双斑螭单位网次渔获量(相对渔获量Y, g/h)。

样品处理和环境因子测定等均按照《海洋调查规范》进行。环境数据由海上同步CTD仪测量获得，包括表层温度(SST)、底层温度(SBT)、表层盐度(SSS)、底层盐度(SBS)、水深(depth)。

### 1.2 模型方法

渔业生态系统中可能存在大量非线性、非可加性过程<sup>[18]</sup>。通常应用物种分布模型模拟生物分布与环境因子的关系。其中GLM模型、GAM模型在渔业领域有着较为广泛的应用<sup>[19]</sup>，其模型结构简单，可以处理非线性和非连续型变量<sup>[20]</sup>。特别是GAM应用非参数的方法检测数

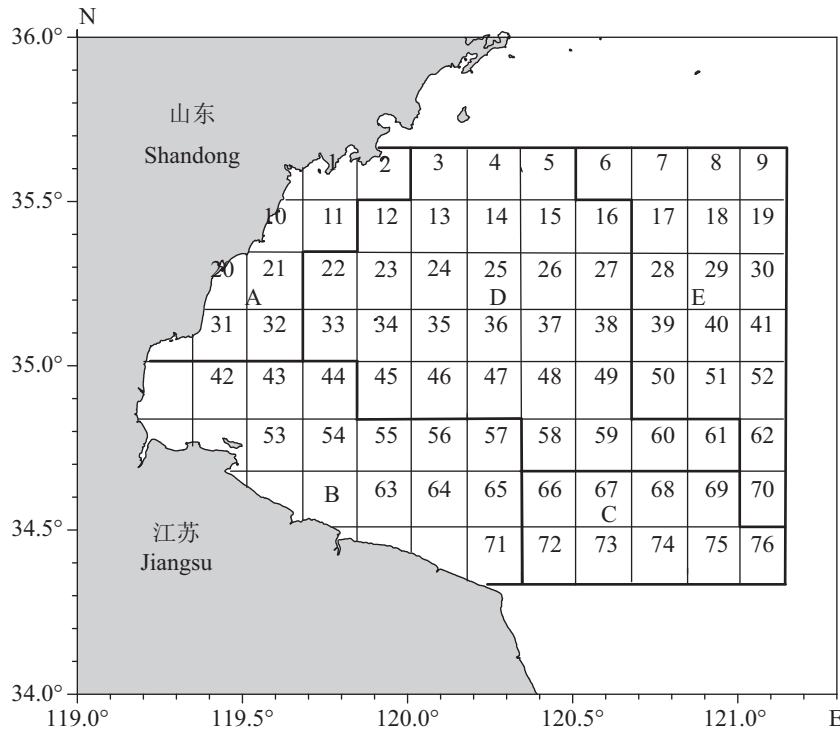


图1 海州湾及邻近海域渔业资源与环境调查区域示意图

Fig. 1 Map of bottom trawl survey areas in Haizhou Bay and adjacent waters

据的结构,更灵活地拟合数据间的复杂关系<sup>[21]</sup>。而随机森林等机器学习模型可以处理数据缺失或数据量大的情况,并具有更少过度拟合的倾向,越来越受到重视。这3种模型均能定量地分析环境因子与资源丰度的关系。因此,本实验利用这3种模型方法定量分析了双斑螭资源分布与环境因子之间的关系,并通过比较不同模型结果的一致性进行相互验证,以对影响双斑螭分布的环境因子做出更准确的判断。

**GLM** GLM是在一般线性回归模型的基础上,通过连接函数对响应变量进行一定的转化,并将指数分布的回归与一般的线性回归相结合,通过最大似然估计的多次迭代推导出各参数值。其回归方程<sup>[22]</sup>:

$$g(Y) = \alpha + \sum_{i=1}^n \beta_i(x_i) + \varepsilon$$

式中, $g()$ 为连接函数, $\alpha$ 为截距项, $\beta_i$ 为 $x_i$ 的回归系数。 $\varepsilon$ 为误差项,与解释变量 $x_i$ 无关,误差分布包括正态、泊松、二项、伽马分布等。

**GAM** GAM是GLM的半参数化扩展,用 $f_i$ 代替回归系数 $\beta_i$ , $f_i()$ 是每个解释变量的非参数函数,假设每个 $f_i$ 均有平滑函数,并通过散点图逐步进行平滑估计,所使用的散点图平滑为立方样条(cubic spline)。回归方程<sup>[23]</sup>:

$$g(Y) = \alpha + \sum_{i=1}^n f_i(x_i) + \varepsilon$$

**随机森林** 随机森林是一种基于分类回归树(CART)算法的集成学习方法<sup>[24]</sup>,它将bagging和随机选择变量结合起来,通过这种方式产生成千上百棵分类回归树,使用投票方法组合最终的结果<sup>[25]</sup>;如果是用于回归,则利用这些回归树预测结果的均值作为最终预测值。

### 1.3 因子筛选与模型拟合

对 $Y$ (相对渔获量)进行对数转换得到 $\ln(Y+1)$ ,使其服从正态分布<sup>[26]</sup>,作为响应变量;选取年(year)、月(month)、经度(longitude)3个时空因子,以及表温、底温、表盐、底盐、深度、离岸距离(distance)6个环境因子,共9个因子作为解释变量。本实验采用方差膨胀因子(VIF)<sup>[27]</sup>对初始9种因子进行多重共线性检验,筛选得到可以加入模型的因子。一般原则下, $\sqrt{VIF} > 2$ 就表明存在多重共线性问题。

模型拟合前,设置GLM和GAM的误差项为高斯分布。拟合过程中,3种模型利用逐步回归法<sup>[28]</sup>进行因子筛选和拟合,即不断添加因子到模型中,运用AIC<sup>[29]</sup>、方差解释率(percentage of variance explained)、残差偏差(residual)等检验模型的拟合效果,AIC值越低,模型拟合效果越好;方差解释率越高,模型拟合效果越好。当模型的AIC值和方差解释率等不再变化,构建过程结束,即得到最适模型。本实验运用卡方检验<sup>[30]</sup>来评估因子对响应变量影响的显著性。

AIC由以下公式计算:

$$AIC = 2k - 2 \ln L$$

式中, $k$ 是参数的个数, $L$ 是似然函数。

### 1.4 模型预测性能评估

为了检验和对比GLM、GAM和随机森林模型的预测效果,本实验采用交叉验证的方法对构建的3个最优模型进行性能评价。即从总数据集中随机选取80%作为训练数据来建立模型,20%作为预测数据来评估模型的性能。计算预测值与观察值的均方根误差RMSE (root mean squared error)<sup>[31]</sup>,将预测值( $P_i$ )与观察值( $O_i$ )进行线性回归得到相应的斜率、截距和决定系数 $R^2$  (coefficient of determination)等性能指标。重复此过程100次。

RMSE由以下公式计算<sup>[31]</sup>:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2}$$

### 1.5 分布预测

模拟环境数据是采用非结构化网格-有限体积方法的FVCOM (finite-volume coastal ocean model)陆架模式对调查站位的实测环境数据进行模拟得到的,此模型已被广泛地应用在沿岸及河口区域进行海岸区域环境预测以及管理<sup>[32-33]</sup>。模拟环境数据包括2011年3月、5月、7月、9月、12月的底层温度、表层盐度、底层盐度、水深,模拟的海域范围为119°~122°E、34°~36°N,分辨率为5'×5',共有1 000网格点。

通过对3个模型比较,选择预测性能最好的模型,将模拟环境数据带入该模型,对2011年海州湾调查水域1 000个均匀覆盖点双斑螭的相对渔获量进行空间补充预测,对比双斑螭资源分布季节间的变化,并运用surfer软件绘制2011

年模拟月份的双斑螭资源分布图。

以上模型构建和检验方法均通过R 3.3.2软件实现(R Core Team, 2017), 其中GAM由“mgcv”包构建, 随机森林模型由“Random Forest”包构建。

## 2 结果

### 2.1 重要因子的筛选

通过方差膨胀因子对包含9个因子的GLM模型进行多重共线性检验, 得到SST、SBT与其他因子存在多重共线性问题, 考虑到双斑螭的底栖特征, 因此舍去SST, 其他因子均可加入模型。表1列出了各模型中逐步加入筛选因子后各指标的变化情况。

GLM最适模型:  $\ln(Y+1)=year+month+depth+SSS+\varepsilon$ ; GLM模型中年份的贡献率最大, 为21.7%; 其次是月份、深度、表层盐度, 分别为8.3%、1.1%和3.1%。(表1)

GAM最适模型:  $\ln(Y+1)=year+month+s(SSS)+s(longitude)+s(SBT)+s(depth)+\varepsilon$ 。其中年份的贡献率最大, 为21.6%; 其次是月份、底温、表层盐度, 分别为8.4%、5.4%和3.6%, 另外经度和深度

也对模型有一定贡献率(表1)。

随机森林最适模型:  $\ln(Y+1)=year+month+depth+longitude$ , 其中年份贡献率最大, 为18.45%; 其次是经度、深度、月份, 对模型的贡献率分别为8.2%、3.9%和3.8%(表1)。

GLM、GAM和随机森林3种模型的最终累积解释率分别为34.2%、42.7%、34.4%, 即GAM对于数据拟合效果较好(表1)。每个模型均有筛选出年份、月份以及深度3个因子。其中, 年份因子是影响双斑螭分布的主要因子, 其次是月份因子, 深度在不同的模型中影响程度不同。

### 2.2 相对渔获量与解释变量的关系

双斑螭的相对渔获量在年间和季节间变化显著; 不同模型中年份对相对渔获量的影响基本一致, 而月份在不同模型中对相对渔获量的影响不同。3个模型均显示2014年和12月分别对应的相对渔获量较多。

在GLM中, 表层盐度和水深对相对渔获量分布均影响显著( $P<0.05$ )(表1); 双斑螭相对渔获量随着表层盐度和水深升高而增大, 并呈线性正相关关系(图2)。在GAM中, 表层盐度、水深

表 1 3种模型拟合结果及各因子重要性

Tab. 1 Results from models fitting and analysis of important factors

模型 model	加入的因子 added factors	累积偏差解释率/% accumulation of deviance explanation	贡献率 importance	卡方检验 $P_t$	AIC
GLM	year	21.7	21.7	\	1 114.9
	+month	30.0	8.3	<0.001***	1 096.4
	+depth	31.1	1.1	<0.05*	1 094.3
	+SSS	34.2	3.1	<0.001***	1 084.8
	year	21.6	21.6	\	1 114.9
GAM	+month	30	8.4	<0.001***	1 096.4
	+s(SSS)	33.6	3.6	<0.001***	1 085.5
	+s(longitude)	35.5	1.9	<0.05*	1 081.7
	+s(SBT)	40.9	5.4	<0.01**	1 071.7
	+s(depth)	42.7	1.8	<0.01**	1 066.3
随机森林 Random Forest	year	18.5	18.5		
	+month	22.3	3.8		
	+depth	26.2	3.9		
	+longitude	34.4	8.2		

注: s表示自然样条平滑; year表示调查年份; month表示调查月份; SBT表示底层海水温度; SSS表示表层海水盐度; \*表示差异显著( $P<0.05$ ), \*\*及\*\*\*表示差异极显著( $P<0.01$ )。随机森林模型本身不提供卡方检验和AIC结果

Notes: s means cubic spline; SBT means the sea bottom temperature; SSS means the sea surface salinity; \* indicates significant difference ( $P<0.05$ ), \*\* and \*\*\* indicate more significant difference ( $P<0.01$ ). The random forest model itself does not provide the chi-square and the AIC value

和底温对相对渔获量分布影响均极显著 ( $P < 0.01$ ) (表1): 相对渔获量随表层盐度和水深的变化趋势与GLM模型基本一致; 底温在5~20 °C时, 相对渔获量呈上升趋势, 在20~25 °C时逐渐下降, 25 °C以后又缓慢增大; 双斑螭相对渔获量随经度的增加先缓慢下降, 到120.5°E之后, 又呈缓慢上升趋势(图3)。随机森林模型中, 相对渔获量随水深增加整体呈上升趋势, 而在达到15 m之前较平缓; 相对渔获量随经度的变化较平滑, 119.5°~120.0°E呈平缓的下降趋势, 120.0°~120.7°E无明显变化, 之后缓慢上升(图4)。3个模型中, 水深、盐度等因素对于双斑螭分布的影响趋势基本一致。

### 2.3 模型的预测性能

通过交叉验证得到100次模型预测结果与观测值的均方根误差(RMSE)、斜率、截距和决定系数 $R^2$ 等性能指标的平均值。随机森林的均方根误差为0.96, 明显小于其他2种模型, 其预测结果与观测值的直线决定系数 $R^2$ 为0.87, 显著高于GAM和GLM, 斜率更接近于1。随机森林模型性能优于GLM和GAM, 预测结果与观测值之间差异更小。GAM和GLM的预测性能相差不大(表2)。

### 2.4 模型预测

通过比较3种模型的预测性能, 发现随机森林模型的预测性能最好, 将FVCOM模拟的环境数据加入随机森林模型中进行预测, 绘制空间分布图。发现5月份双斑螭丰度分布呈现出东部海域高, 西部近岸海域低的特点; 其他月份双斑螭的相对渔获量分布规律大体一致, 均为西南部水域低于中北部(图5)。

## 3 讨论

### 3.1 3种模型的比较

3种模型中, GAM对本实验数据集拟合效果最好; 随机森林模型对双斑螭相对渔获量的预测效果最佳。该结论说明模型对现有数据的拟合效果不能保证模型预测性能。GAM拟合过程中, 比GLM能解释更多变量间的非线性关系, 使GAM中包含了更多的解释变量从而提高了偏差解释率。同时, 拟合结果表明GAM比GLM对环境因子的变化更加敏感, 能够解释更多空间数据的变化, 这与倪一卓等<sup>[15]</sup>研究结论一致。另外, 本实验运用AIC准则对GLM和GAM进行模型选择, 而AIC准则在样本量较少的情况下, 可能

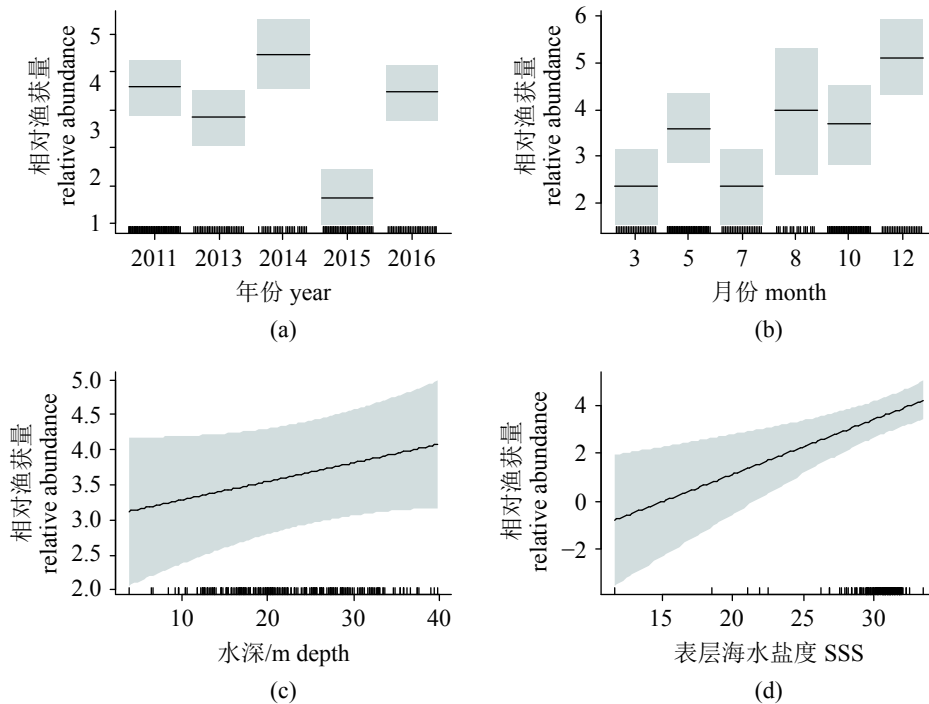


图 2 各影响因子与双斑螭相对渔获量之间关系的GLM分析图  
 Fig. 2 Effects of factors on *C. bimaculata*'s relative abundance in GLM

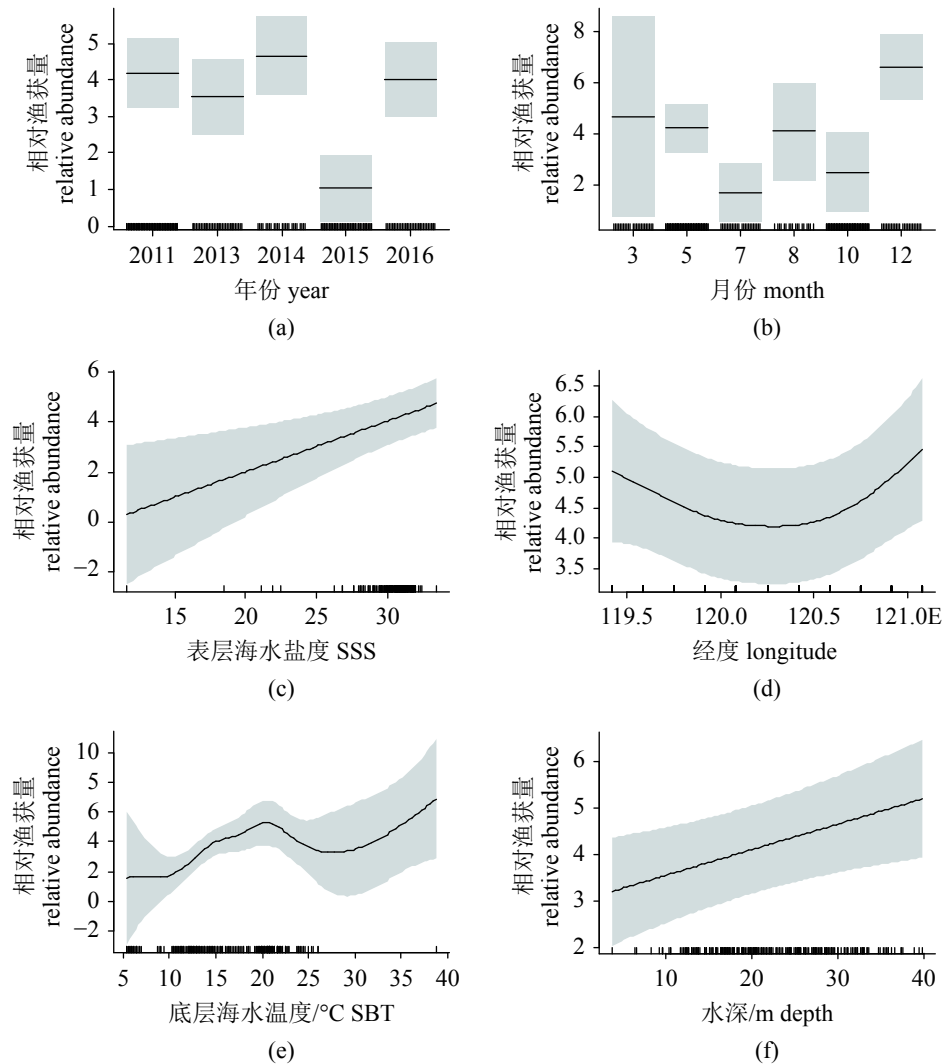


图3 各影响因子与双斑蟳相对渔获量之间关系的GAM分析图

Fig. 3 Effects of factors on *C. bimaculata*'s relative abundance in GAM

会高估参数数量的影响,使得GAM和GLM倾向于选择较复杂的模型<sup>[34]</sup>。另一方面,本实验结果显示,GAM的拟合效果好而预测能力较差,可能反映了GAM对数据存在过度拟合的现象,即对样本噪声进行了解释,而偏离了对真实值的解释,使得仅对训练数据有较好的拟合效果,对训练数据集之外的预测能力较差。随机森林模型具有集成学习的特点,即通过聚集多个回归树的结果来提高预测精度<sup>[35]</sup>,构造不同的训练集增加回归树间的差异,提高外推预测能力,但同时可能降低了其对训练数据集的拟合效果<sup>[36]</sup>。此外,随机森林模型能识别因子间的交互效应<sup>[37]</sup>,可以处理数据点相对于解释变量较少的情况<sup>[38]</sup>。GLM、GAM与随机森林等集成学习方法相比,缺少对数据集进行多次的重复抽样拟

合,此外本实验为了避免无法拟合和便于模型解释,在GLM和GAM没有考虑解释变量间交互效应,可能使模型没有充分解释相对渔获量与因子的关系<sup>[16, 20]</sup>;而随机森林模型则基于回归树计算,避免了该限制。

### 3.2 模型解释变量对双斑蟳分布的影响

在3种模型筛选的所有因子中,年份因子对双斑蟳相对渔获量影响最大,即双斑蟳相对渔获量在年间有极大的变异性。这可能是由于双斑蟳的生命周期短(仅为1年)<sup>[7]</sup>,其资源量易受多种环境因素的影响,如水温、盐度等环境因子波动、捕食者数量变化、特别是其补充群体的数量波动,均会对双斑蟳相对渔获量产生影响。月份因子在3种模型中影响仅次于年份,说

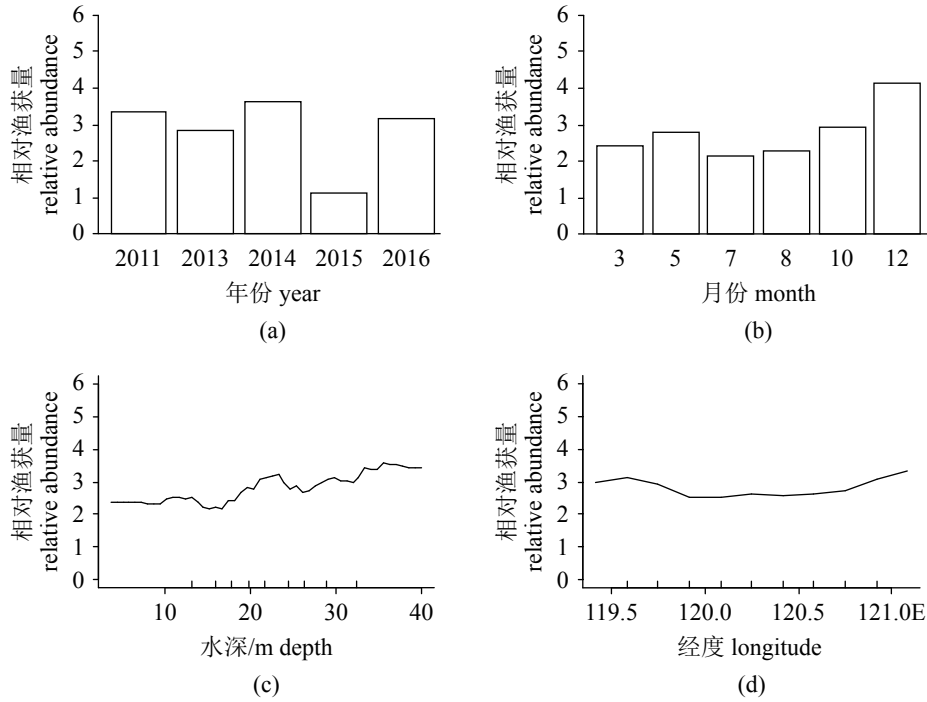


图 4 各影响因子与双斑螭相对渔获量之间关系的随机森林模型分析图

Fig. 4 Effects of factors on *C. bimaculata*'s relative abundance in Random Forest

表 2 3种模型交叉验证结果

Tab. 2 Cross-validation comparison between three models

模型 model	均方根误差 root mean squared error	斜率 slope	截距 intercept	决定系数R <sup>2</sup> coefficient of determination
GAM	2.08	0.35	1.78	0.30
GLM	1.97	0.34	1.85	0.34
随机森林 Random Forest	0.96	0.70	0.87	0.87

明双斑螭相对渔获量分布在不同月份间变化明显。月份作为时间因子,对双斑螭丰度的影响是环境理化因子等综合作用的结果,由于海州湾水域受到不同季节的海流和水团的影响,以及季节间温度的变化等,使得双斑螭相对渔获量出现季节间差异。3种模型均显示12月份双斑螭相对渔获量高于其他月份,可能是因为其繁殖高峰期在夏、秋季<sup>[1, 8]</sup>,冬季时的补充群体数量明显增多,并且补充群体增加的生物量大于死亡群体减少的生物量,因此其种群生物量增加。

在因子筛选过程中,3种模型均认为水深对双斑螭资源分布影响显著,且与双斑螭丰度大致呈正相关关系。尤其是在夏季调查航次的双斑螭相对渔获量分布图中,海州湾西南部站位均未捕获到双斑螭,在中部和东北部站位相对

渔获量较大(图6),这与海州湾水域东北部高、西南部低的水深分布特点相吻合<sup>[39]</sup>,夏季的分布特征极大地促进了水深被筛选为重要影响因素。需要注意的是,水深作为一个综合性的环境因子,不是直接影响双斑螭相对渔获量分布,而是通过温度、盐度、溶解氧等其他环境因子影响双斑螭分布<sup>[40]</sup>。因此,西南部苏北浅滩(<20 m)无双斑螭的分布,说明此水域的某些生境特征可能不适于双斑螭的栖息,在今后的研究中需要进一步探索。另外,GLM和GAM均表明SSS对双斑螭丰度影响极显著( $P < 0.001$ ),并且高于其他环境因子,但表层盐度对其影响机制并不明确。表层盐度的变化可能表征了海州湾附近海流和水团的季节性变化,而复杂的海流变化会对双斑螭分布产生影响,因此表层盐度可能间接影响了其分布。双斑螭生态习性与底层生境有关,其食性与其他小型底栖蟹类相似,以底栖甲壳类、双壳类以及碎屑藻类等为主,这些饵料生物也易受到底层生境的影响,也有研究表明,双斑螭在粉砂质粘土类型的底质中资源密度最高<sup>[41-43]</sup>。本实验将底层温度和底质类型(sediment)加入模型中:底层温度只在GAM模型中对双斑螭相对渔获量影响显著,并呈曲线变

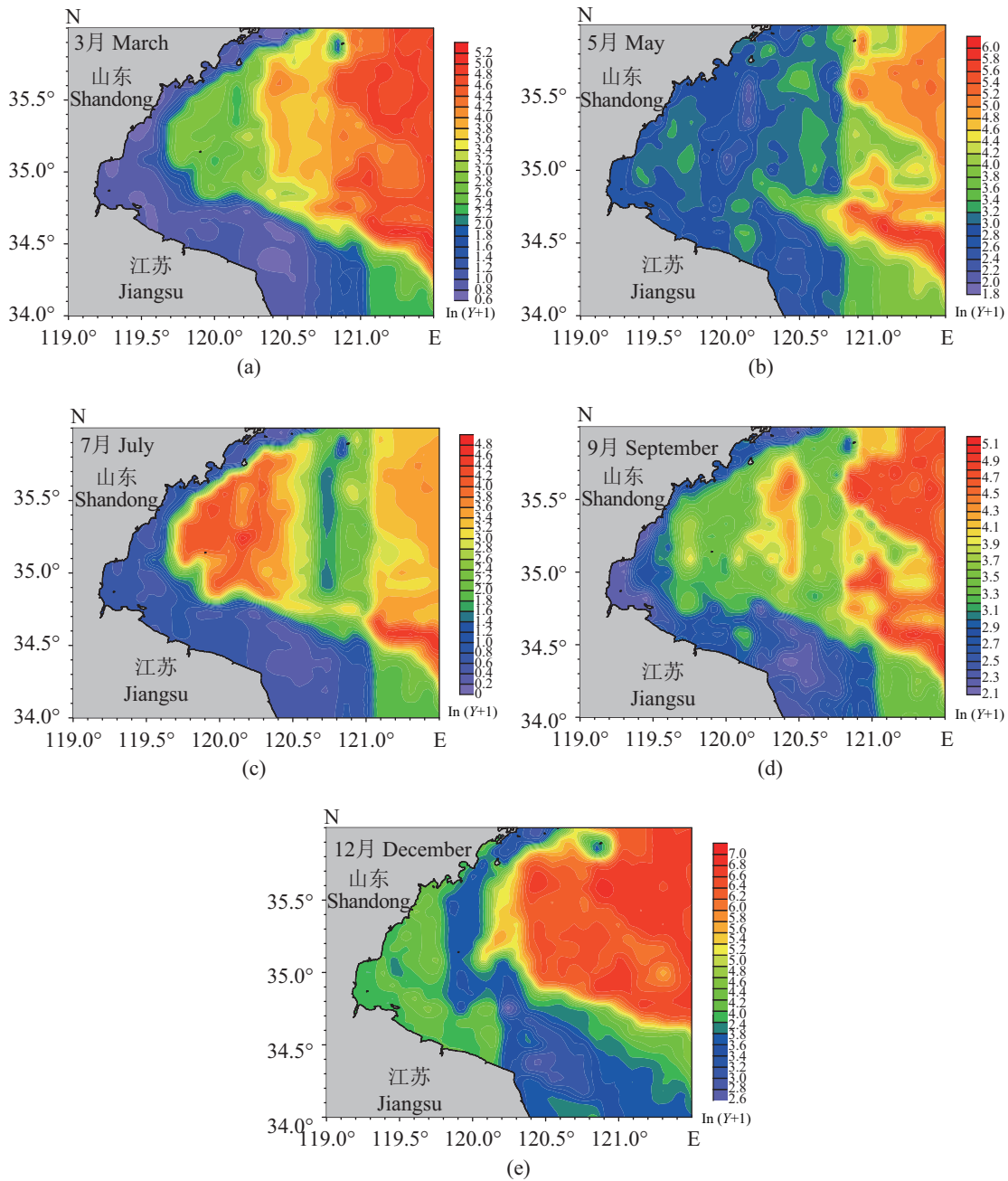


图 5 双斑螭不同季节相对渔获量空间预测分布图

Fig. 5 Predicted spatial *C. bimaculata*'s relative abundance distribution in different seasons

化, 这种变化不符合生态学意义, 由于本实验中检验到底温与其他因子(如月份、水深等)之间存在一定的共线性, 会造成模型拟合结果不稳定, 并且影响底温对残差的解释程度, 从而使得底温与双斑螭相对渔获量关系曲线波动性较大; 在3种模型中, 底质类型对双斑螭相对渔获量分布均无显著影响。可能的原因, 一方面底质类型并不是影响双斑螭分布的重要因素; 另一方面, 海州湾调查区域的底质类型相差不

大, 中部、远岸海域及西南海域均以粉砂为主<sup>[39]</sup>, 仅有西北近岸1~2个调查站位附近属于礁岩底质, 因此对双斑螭分布无明显影响。

利用FVCOM模拟的环境数据在随机森林中对2011年各季节双斑螭资源分布进行预测, 结果显示3月和12月的分布图中双斑螭在海州湾外海的相对渔获量较高, 并沿近岸逐渐降低(图5)。这与冬季海州湾黄海暖水舌分支由东向西进入海州湾中部远岸水域的位置基本相符合, 可能



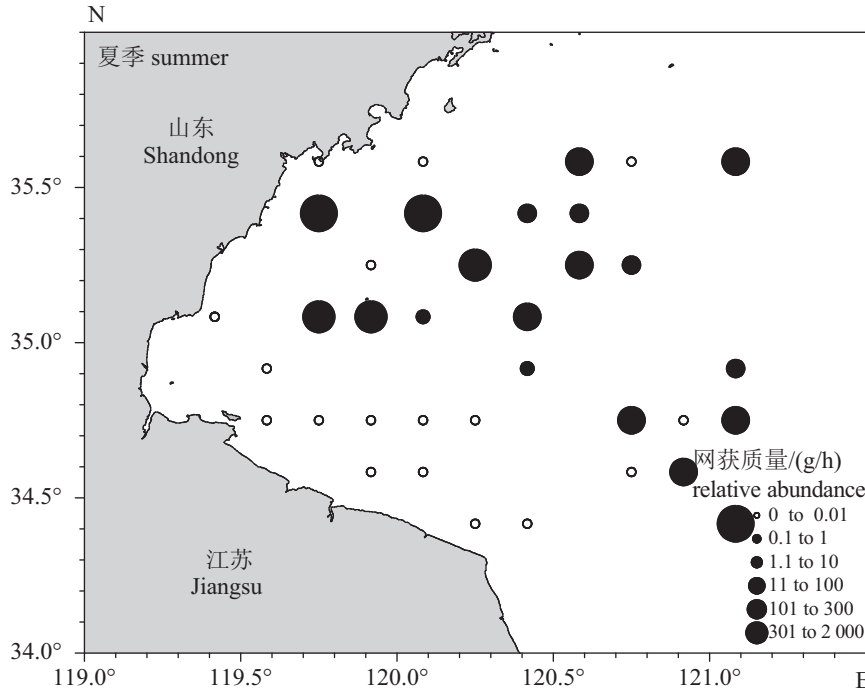


图 6 双斑螭夏季调查航次的相对渔获量分布图

Fig. 6 The spatial distribution of the relative abundance of *C. bimaculata* in summer

说明黄海暖流对其分布有一定的影响；但是黄海暖流具有低溶解氧、高温高盐的特征<sup>[44]</sup>，而在日本中部伊势湾的研究中则发现低溶解氧会造成双斑螭丰度的下降<sup>[8]</sup>，这与低溶解氧的黄海暖流下双斑螭相对渔获量较高的结论相反，可能是由于不同研究中不同海域溶解氧的水平情况不同，而黄海暖流带来的低溶解氧水平未对双斑螭丰度产生影响。另外，2011年双斑螭相对渔获量的空间预测分布除5月之外，均与海州湾水深分布特点相似，说明水深作为主要环境因子在随机森林模型中影响显著，并且影响模型预测结果。

#### 4 结论

本实验运用3种模型，分析了双斑螭资源量分布与重要的环境、时空因子之间的关系，通过3种模型结果的相互验证，得到重要影响因子有年份、月份、表层盐度和水深，并且在3种模型中这些因子对双斑螭分布的影响趋势基本一致，说明该结果具有一定的可靠性。GLM、GAM模型主要是基于统计学方法，而随机森林模型主要基于机器学习方法，这3种模型均没有直接涉及生态学过程，模型的解释更多地依赖于对双斑螭生活史特征的现有认识。随机森林

模型的预测效果最好，GAM则展现了最好的拟合效果，这对于今后模型的选择有一定参考价值；需要注意实际中最适合的SDM可能因研究对象、调查方法、调查范围等的差异而不同<sup>[21]</sup>，在模型比较和选择中应谨慎。

本实验由于双斑螭资源调查受限于空间范围和随机采样点的数量，模型分析结果可能受到一定程度的影响，而不能完整地反映双斑螭的物种分布情况<sup>[12]</sup>。此外，由于数据量较少，本实验在GLM和GAM模型中加入一些交互项时，由于对应的自由度明显增大，显示无法实现模型的拟合。今后的研究应在更大空间范围和分辨率更高的取样数据下进行，并充分考虑变量间的交互效应以提高模型的性能。本实验中3个模型对双斑螭资源分布的解释率并不高，说明现有的解释变量不能完全反映该物种栖息地的生境，可能存在其他与响应变量关系密切的环境因子，而这些环境因子没有进行观测。因此在今后的调查中应补充观测更多的环境因子，以便可以最大限度地解释响应变量。本实验3个模型结果均表明时间因子的效应大于环境因子，下一步的研究可以进行解释变量筛选，找出具有时间特征的环境因子来替代时间因子，使得模型的解释更能反映资源分布与环境因子

的关系<sup>[45]</sup>。另外,环境因子影响其分布的具体生态学机制有待进一步分析。

### 参考文献:

- [1] 郑伟. 东、黄海区日本螯和双斑螯的分子系统地理学研究[D]. 舟山: 浙江海洋学院, 2015.  
Zheng W. Molecular phylogeography study of *Charybdis japonica* and *Charybdis bimaculata* in East China sea and Yellow sea[D]. Zhoushan: Zhejiang Ocean University, 2015 (in Chinese).
- [2] 潘国良, 朱增军, 张洪亮, 等. 浙南沿岸产卵场春季双斑螯生物量分布及其与环境因子的关系[J]. 浙江海洋学院学报(自然科学版), 2012, 31(6): 482-486.  
Pan G L, Zhu Z J, Zhang H L, et al. Distribution of the biomass of *Charybdis bimaculata* and its relationships with the environmental factors in the coastal spawning ground of South Zhejiang during spring[J]. Journal of Zhejiang Ocean University (Natural Science Edition), 2012, 31(6): 482-486(in Chinese).
- [3] 逢志伟. 胶州湾虾蟹类群落结构及主要种类渔业生物学特征[D]. 青岛: 中国海洋大学, 2014.  
Pang Z W. Community structure of the shrimp and crab assemblages and fishery biological characteristics of some important species in the Jiaozhou Bay[D]. Qingdao: Ocean University of China, 2014 (in Chinese).
- [4] 张波, 唐启升, 金显仕. 黄海生态系统高营养层次生物群落功能群及其主要种类[J]. 生态学报, 2009, 29(3): 1099-1111.  
Zhang B, Tang Q S, Jin X S. Functional groups of communities and their major species at high trophic level in the Yellow Sea ecosystem[J]. Acta Ecologica Sinica, 2009, 29(3): 1099-1111(in Chinese).
- [5] 李忠义, 左涛, 戴芳群, 等. 长江口及南黄海水域春季生物摄食生态的稳定同位素研究[J]. 水产学报, 2009, 33(5): 784-789.  
Li Z Y, Zuo T, Dai F Q, et al. Study on feeding habits of organisms from Changjiang Estuary and adjacent Southern Yellow Sea in spring with stable isotope technology[J]. Journal of Fisheries of China, 2009, 33(5): 784-789(in Chinese).
- [6] 麻秋云, 韩东燕, 刘贺, 等. 应用稳定同位素技术构建胶州湾食物网的连续营养谱[J]. 生态学报, 2015, 35(21): 7207-7218.  
Ma Q Y, Han D Y, Liu H, et al. Construction of a continuous trophic spectrum for the food web in Jiaozhou Bay using stable isotope analyses[J]. Acta Ecologica Sinica, 2015, 35(21): 7207-7218(in Chinese).
- [7] Doi W, Yokota M, Strüssmann C A, et al. Growth and reproduction of the portunid crab *Charybdis bimaculata* (Decapoda: Brachyura) in Tokyo Bay[J]. Journal of Crustacean Biology, 2008, 28(4): 641-651.
- [8] Narita T, Ganmanee M, Sekiguchi H. Population dynamics of portunid crab *Charybdis bimaculata* in Ise Bay, central Japan[J]. Fisheries Science, 2008, 74(1): 28-40.
- [9] Pyke C R. Habitat loss confounds climate change impacts[J]. Frontiers in Ecology and the Environment, 2004, 2(4): 178-182.
- [10] Lotze H K, Lenihan H S, Bourque B J, et al. Depletion, degradation, and recovery potential of estuaries and coastal seas[J]. Science, 2006, 312(5781): 1806-1809.
- [11] 朱国平, 朱小艳, 徐怡瑛, 等. 基于GAM模型分析夏秋季南奥克尼群岛南极磷虾渔场时空分布及与环境因子之间的关系[J]. 极地研究, 2012, 24(3): 266-273.  
Zhu G P, Zhu X Y, Xu Y Y, et al. The spatiotemporal distribution of fishing grounds for Antarctic krill (*Euphausia superba*) around the South Orkney Islands in austral summer-autumn and its relation to environmental factors based on a generalized additive model[J]. Chinese Journal of Polar Research, 2012, 24(3): 266-273(in Chinese).
- [12] 许仲林, 彭焕华, 彭守璋. 物种分布模型的发展及评价方法[J]. 生态学报, 2015, 35(2): 557-567.  
Xu Z L, Peng H H, Peng S Z. The development and evaluation of species distribution models[J]. Acta Ecologica Sinica, 2015, 35(2): 557-567(in Chinese).
- [13] 陈雪忠, 樊伟, 崔雪森, 等. 基于随机森林的印度洋长鳍金枪鱼渔场预报[J]. 海洋学报, 2013, 35(1): 158-164.  
Chen X Z, Fan W, Cui X S, et al. Fishing ground forecasting of *Thunnus alalung* in Indian Ocean based on random forest[J]. Acta Oceanologica Sinica, 2013, 35(1): 158-164(in Chinese).
- [14] 杨胜龙, 张禹, 张衡, 等. 不同模型在渔业CPUE标准化中的比较分析[J]. 农业工程学报, 2015, 31(21): 259-264.  
Yang S L, Zhang Y, Zhang H, et al. Comparison and analysis of different model algorithms for CPUE standardization in fishery[J]. Transactions of the Chinese Society of Agricultural Engineering, 2015, 31(21): 259-264(in Chinese).
- [15] 倪一卓, 程和琴, 江红, 等. 鱼类栖息地模拟的比较研

- 究—以东海鲈鱼为例[J]. 水产科学, 2009, 28(12): 726-732.
- Ni Y Z, Cheng H Q, Jiang H, *et al.* Fish habitat modeling in the East China Sea: a case of chub mackerel[J]. Fisheries Science, 2009, 28(12): 726-732(in Chinese).
- [16] Campbell R A. CPUE standardisation and the construction of indices of stock abundance in a spatially varying fishery using general linear models[J]. Fisheries Research, 2004, 70(2-3): 209-227.
- [17] 官文江, 陈新军, 高峰, 等. GLM模型和回归树模型在CPUE标准化中的比较分析[J]. 上海海洋大学学报, 2014, 23(1): 123-130.
- Guan W J, Chen X J, Gao F, *et al.* Comparisons of regression tree and GLM performance in CPUE standardization[J]. Journal of Shanghai Ocean University, 2014, 23(1): 123-130(in Chinese).
- [18] 官文江, 陈新军, 高峰, 等. 海洋环境对东、黄海鲈鱼灯光围网捕捞效率的影响[J]. 中国水产科学, 2009, 16(6): 949-958.
- Guan W J, Chen X J, Gao F, *et al.* Environmental effects on fishing efficiency of *Scomber japonicus* for Chinese large lighting purse seine fishery in the Yellow and East China Seas[J]. Journal of Fishery Sciences of China, 2009, 16(6): 949-958(in Chinese).
- [19] Yu H, Jiao Y, Carstensen L W. Performance comparison between spatial interpolation and GLM/GAM in estimating relative abundance indices through a simulation study[J]. Fisheries Research, 2013, 147: 186-195.
- [20] Guisan A, Edwards Jr T C, Hastie T. Generalized linear and generalized additive models in studies of species distributions: setting the scene[J]. Ecological Modelling, 2002, 157(2-3): 89-100.
- [21] 曹铭昌, 周广胜, 翁厚生. 广义模型及分类回归树在物种分布模拟中的应用与比较[J]. 生态学报, 2005, 25(8): 2031-2040.
- Cao M C, Zhou G S, Weng E S. Application and comparison of generalized models and classification and regression tree in simulating tree species distribution[J]. Acta Ecologica Sinica, 2005, 25(8): 2031-2040(in Chinese).
- [22] 陈新军, 高峰, 官文江, 等. 渔情预报技术及模型研究进展[J]. 水产学报, 2013, 37(8): 1270-1280.
- Chen X J, Gao F, Guan W J, *et al.* Review of fishery forecasting technology and its models[J]. Journal of Fisheries of China, 2013, 37(8): 1270-1280(in Chinese).
- [23] Hastie T J, Tibshirani R J. Generalized Additive Models[M]. London, New York: Chapman and Hall, 1990.
- [24] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [25] Lantz B. Machine Learning with R[M]. 2nd ed. Birmingham: Packt Publishing, 2015.
- [26] Beverton R J H, Holt S J. On the Dynamics of Exploited Fish Populations[M]. Dordrecht: Springer, 1993.
- [27] Kabacoff R I. R in Action: Data Analysis and Graphics with R[M]. Greenwich: Manning Publications, 2011: 1-474.
- [28] Wood S N. Stable and efficient multiple smoothing parameter estimation for generalized additive models[J]. Journal of the American Statistical Association, 2004, 99(467): 673-686.
- [29] Burnham K P, Anderson D R. Model selection and inference: A practical Information-theoretic approach[M]. New York: Springer, 2002: 175-196.
- [30] Scheipl F, Greven S, Küchenhoff H. Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models[J]. Computational Statistics & Data Analysis, 2008, 52(7): 3283-3299.
- [31] Hyndman R J, Koehler A B. Another look at measures of forecast accuracy[J]. International Journal of Forecasting, 2006, 22(4): 679-688.
- [32] Xue Y, Guan L S, Tanaka K, *et al.* Evaluating effects of rescaling and weighting data on habitat suitability modeling[J]. Fisheries Research, 2017, 188: 84-94.
- [33] Chen C S, Beardsley R C, Cowles G. An unstructured grid, finite-volume coastal ocean model (FVCOM) system[J]. Oceanography, 2006, 19(1): 78-89.
- [34] Shono H. Is model selection using Akaike's information criterion appropriate for catch per unit effort standardization in large samples?[J]. Fisheries Science, 2005, 71(5): 978-986.
- [35] 蔡林霖. 随机森林的模型选择及其并行化方法[D]. 哈尔滨: 哈尔滨工业大学, 2012.
- Cai L L. Model selection of random forest and its parallelization[D]. Harbin: Harbin Institute of Technology, 2012 (in Chinese).
- [36] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, 26(3): 32-38.
- Fang K N, Wu J B, Zhu J P, *et al.* A review of technolo-

- gies on random forests[J]. *Statistics & Information Forum*, 2011, 26(3): 32-38(in Chinese).
- [37] Li Z G, Ye Z J, Wan R, *et al.* Model selection between traditional and popular methods for standardizing catch rates of target species: a case study of Japanese Spanish mackerel in the gillnet fishery[J]. *Fisheries Research*, 2015, 161: 312-319.
- [38] Strobl C, Boulesteix A L, Zeileis A, *et al.* Bias in random forest variable importance measures: illustrations, sources and a solution[J]. *BMC Bioinformatics*, 2007, 8: 25.
- [39] 苏巍. 海州湾海域鱼类群落多样性及其与环境因子的关系[D]. 青岛: 中国海洋大学, 2014.  
Su W. Fish community diversity in Haizhou Bay and its relationship with environmental factors[D]. Qingdao: Ocean University of China, 2014 (in Chinese).
- [40] 邱永松. 南海北部大陆架鱼类群落的区域性变化[J]. *水产学报*, 1988, 12(4): 303-313.  
Qiu Y S. The regional changes of fish community on the northern continental shelf of South China Sea[J]. *Journal of Fisheries of China*, 1988, 12(4): 303-313(in Chinese).
- [41] 蒋日进. 枸杞岛近岸海域食物网结构研究[D]. 上海: 上海海洋大学, 2015.  
Jiang R J. Research on the food web structure in the costal areas of Gouqi Island[D]. Shanghai: Shanghai Ocean University, 2015 (in Chinese).
- [42] 黄美珍. 台湾海峡及邻近海域主要无脊椎动物食性特征及其食物关系研究[J]. *海洋科学*, 2005, 29(1): 73-80.  
Huang M Z. Study on feeding habits and food relationship of main economic invertebrates in Taiwan Strait and its adjacent areas[J]. *Marine Sciences*, 2005, 29(1): 73-80(in Chinese).
- [43] 赵静, 刘涵, 原振政, 等. 日本蟳对3种贝类的摄食选择及摄食节律的研究[J]. *大连海洋大学学报*, 2012, 27(3): 226-230.  
Zhao J, Liu H, Yuan Z Z, *et al.* Selective feeding on three bivalves and feeding rhythm in Asian swimming crab *Charybdis japonica*[J]. *Journal of Dalian Fisheries University*, 2012, 27(3): 226-230(in Chinese).
- [44] 宋新, 林霄沛, 王悦. 冬季黄海暖流变化及其原因浅析[J]. *中国海洋大学学报(自然科学版)*, 2009, 39(S1): 259-266.  
Song X, Lin X P, Wang Y. The variability of the Yellow sea warm current axis in winter and its possible reason[J]. *Periodical of Ocean University of China(Natural Science Edition)*, 2009, 39(S1): 259-266(in Chinese).
- [45] Bradter U, Kunin W E, Altringham J D, *et al.* Identifying appropriate spatial scales of predictors in species distribution models with the random forest algorithm[J]. *Methods in Ecology and Evolution*, 2013, 4(2): 167-174.

## Relationship between catch distribution of Portunid crab (*Charybdis bimaculata*) and environmental factors based on three species distribution models in Haizhou Bay

LUAN Jing<sup>1</sup>, ZHANG Chongliang<sup>1\*</sup>, XU Binduo<sup>1</sup>, XUE Ying<sup>1</sup>, REN Yiping<sup>1,2</sup>

(1. College of Fisheries, Ocean University of China, Qingdao 266003, China;

2. Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China)

**Abstract:** There are internal relationships between the spatial and temporal distribution of species and environmental factors; however, the randomness and uncertainty of marine ecosystems prevent demonstrating the relationships in an easy way. Based on the seasonal bottom trawl survey data collected from 2011 to 2016 in Haizhou Bay and its adjacent waters, we used three species distribution models, including GLM, GAM and Random Forest to study the spatial distribution of Portunid crab (*Charybdis bimaculata*) and analyze the relation with environmental factors. We selected influential environmental variables and built models according to criteria such as AIC, deviance explanation and cross validation. The effects of environmental variables on the distribution of *C. bimaculata* were evaluated on the basis of the SDMs. The results showed similar results among the three models in the interpretation of the relationship between environmental variables and crab distribution. The GAM provided better fit to data, whereas random forest have superior predictive performances than the other models. The models illustrated significant variations of crab abundance among years and seasons, which contributed to over 18 percent and 3.8 percent of total deviance, respectively. Depth and sea surface salinity were influential environmental factors for relative abundance, both positive correlated with relative catches. The spatial distribution of Portunid crab was featured by high catches in northeast area and low in southwest, in accordance with the water depth in Haizhou Bay. We predicted the spatial distribution of *C. bimaculata* across seasons using FVCOM simulation data and the random forest model to facilitate the exploitation and conservation of the fishery resources.

**Key words:** *Charybdis bimaculata*; GLM; GAM; Random Forest; model comparison; cross validation

**Corresponding author:** ZHANG Chongliang. E-mail: zhangclg@ouc.edu.cn

**Funding projects:** Fundamental Research Funds for the Central Universities (201612002,201712015)