



## 应用 STAR 模型研究海州湾小黄鱼春季资源量的时空分布

赵伟<sup>1,2</sup>, 任一平<sup>1,2,3</sup>, 徐宾铎<sup>1,2</sup>, 薛莹<sup>1,2</sup>, 张崇良<sup>1,2\*</sup>

(1. 中国海洋大学水产学院, 山东 青岛 266003;

2. 海州湾渔业生态系统教育部野外科学观测研究站, 山东 青岛 266003;

3. 青岛海洋科学与技术试点国家实验室, 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:** 根据 2011 年、2013—2016 年春季在海州湾进行的渔业资源调查数据, 应用结构化加性回归 (structured additive regression, STAR) 模型框架, 结合 delta 方法, 根据对空间数据的不同处理方式构建了 5 种物种分布模型, 并比较各模型对出现概率和资源量 2 种数据类型的拟合效果、残差空间独立性和预测性能。结果显示, 加入空间项后模型拟合效果提升, 残差空间自相关性显著降低, 且正态模型和 delta 模型的提升较二项模型明显。空间加性模型 (geoadditive models) 的 AIC 值在二项模型和正态模型中均为最低, 较无空间项广义可加模型 (generalized additive model, GAM) 分别下降 7.60 和 144.90。模型拟合上, 变系数模型 (varying coefficient models, VCM) 的决定系数和 AUC 均最高, 分别为 0.68 和 0.94。预测性能上, 空间加性模型交叉验证的 AUC 值为 (0.793±0.100) 最高, 均方根误差 (RMSE) 值为 (21.65±4.83) 最低, 表明对小黄鱼出现概率和资源密度的估计均最准确。在最优模型的基础上, 根据无结构网格、有限体积、自由表面三维原始方程的海洋环流模型 (FVCOM) 模拟环境数据, 利用 delta 空间加性模型预测海州湾小黄鱼春季资源的时空分布。研究表明, 小黄鱼资源分布主要集中于海州湾南部和东部近岸水域 (34.0°N~34.5°N, 121.0°E~121.5°E), 随着水深的增加而逐渐减少, 且年间变动明显。本研究旨在为海州湾小黄鱼渔业资源的开发和保护提供科学依据。

**关键词:** 小黄鱼; 结构化加性回归; 空间效应; 交叉验证; 模型比较; 海州湾

**中图分类号:** S 932.4

**文献标志码:** A

物种分布模型 (species distribution models, SDMs) 是研究气候变化对物种潜在分布影响的重要方法, 能够为海洋渔业资源空间规划和管理提供重要决策和支持<sup>[1-2]</sup>。其中广义线性模型 (generalized linear model, GLM) 与广义可加模型 (generalized additive model, GAM) 在渔业领域应用广泛, 常被用来分析渔业资源数量分布与环境因子之间的关系和预测渔业资源的时空分布<sup>[3-5]</sup>。但由于上述模型在结构方面存在局限性, 难以处理渔业资源

的空间分布格局和时空交互作用<sup>[6]</sup>。结构化加性回归 (structured additive regression, STAR) 模型是对传统 GAM 的拓展, 为处理变量之间的非线性关系提供了更灵活的框架和函数形式, 向 GAM 加入空间效应、变系数效应和随机效应等, 拓展为空间加性模型 (geoadditive models)、变系数模型 (varying coefficient models, VCM) 和广义加性混合模型 (generalized additive mixed models, GAMM) 等<sup>[7]</sup>。

收稿日期: 2020-09-19 修回日期: 2021-01-07

资助项目: 山东省支持青岛海洋科学与技术试点国家实验室重大科技专项 (2018SDKJ0501-2); 国家重点研发计划 (2018YFD0900904); 国家自然科学基金 (3172852)

第一作者: 赵伟 (照片), 从事鱼类空间分布研究, E-mail: zhaoweiy9346@stu.ouc.edu.cn

通信作者: 张崇良, 从事渔业资源评估研究, E-mail: zhangclg@ouc.edu.cn



渔业资源调查数据中常常占有较高比例的零值, 如果建模过程中零膨胀问题没有被合理解决, 将会使模型的独立性假设无效, 影响其参数的估计并降低模型预测能力<sup>[8]</sup>。针对不同的数据类型, 如在因变量为离散型数据类型(如尾数等)时, 常采用零膨胀泊松分布等, 而针对连续性数据(如渔获质量等)类型, delta 方法是目前广泛采用的方法, 其假设存在两个独立过程决定渔获率和渔获量, 最终实际观测到的渔获量取决于这两个过程<sup>[9]</sup>。

小黄鱼 (*Larimichthys polyactis*) 隶属鲈形目 (Perciformes) 石首鱼科 (Sciaenidae) 黄鱼属 (*Larimichthys*), 系暖温性底层鱼类, 广泛分布于黄海、渤海、东海及朝鲜半岛西岸海域<sup>[10-11]</sup>。近年来, 受捕捞压力和气候变化的影响, 小黄鱼的种群数量及其分布特征发生较大的变化, 渔获物呈现小型化、低龄化、性成熟提前等特点<sup>[12]</sup>。为促进其资源的可持续利用, 诸多学者针对不同海域小黄鱼时空分布开展了调查研究。如刘尊雷等<sup>[13]</sup>集成多种物种分布模型研究了黄海南部小黄鱼冬季分布, 刘梦影等<sup>[14]</sup>应用栖息地适宜性指数模型研究了温台渔场小黄鱼的空间分布特征, 尹洁等<sup>[15]</sup>应用 two-stage GAM 研究海州湾及其邻近海域小黄鱼鱼卵的时空分布特征, 邹易阳等<sup>[16]</sup>应用栖息地适宜性指数模型研究海州湾小黄鱼的空间分布特征。然而, 这些研究较少考虑小黄鱼的空间分布结构和自相关特征及调查数据中零值较多的问题。

本实验根据 2011、2013—2016 年春季在海州湾进行的底拖网调查数据, 比较了 5 种 STAR 模型对海州湾小黄鱼春季资源时空分布的预测性能。此外, 针对数据零值较多的情况, 采用了 delta 方法构建模型, 选择最佳方案对海州湾小黄鱼资源的时空分布进行预测, 以期为合理预测海州湾小黄鱼资源分布的时空变动并为其渔业生产和可持续利用提供科学依据。

## 1 材料与方法

### 1.1 数据来源

数据来源于 2011 年、2013—2016 年春季(5 月)在海州湾及其邻近海域开展的底拖网调查。调查海域范围为 34°20'N~35°40'N, 119°20'E~121°10'E。调查站位设计采用分层随机取样<sup>[17]</sup>, 根据调查海域的水深和纬度等因素的差异, 将调查海域分为 A~E 5 个区域(图 1), 每个航次在各区域内随机选

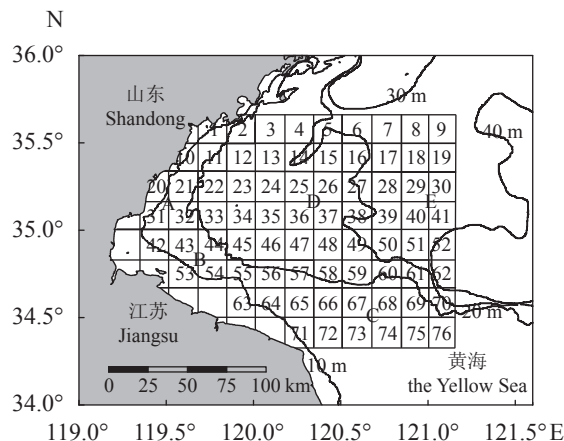


图 1 海州湾及其邻近海域调查站位示意图

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

取一定数量的站位调查, 其中 2011 年调查站位数为 24 个, 其他年份为 18 个。

调查所采用的单拖渔船的功率为 220 kw, 拖速为 2~3 kn, 每站拖网时间约为 1 h, 使用温盐深仪同步采集各个站位的理化环境数据。调查网具网口宽度 25 m, 网口高度约 6 m, 囊网网目尺寸 17 mm。渔获样品带回实验室统计渔获量, 并进行拖网时间(1 h)以及拖速(2 kn)的标准化处理。样品的采集和分析均按照最新的《海洋调查规范》进行<sup>[18]</sup>。

### 1.2 模型方法

选取小黄鱼的资源量 ( $Y$ , kg/km<sup>2</sup>) 作为响应变量, 结合以往对小黄鱼空间分布的相关研究内容<sup>[16,19]</sup>, 选取底层海水温度 (SBT, °C)、底层海水盐度 (SBS)、水深 (Dep, m) 3 个环境因子和站位经度、纬度和年份 3 个时空因子作为初始解释变量。利用 STAR 模型框架, 采用不同方式处理空间位置变量, 构建 5 种模型 (M1~M5):

$$M1: Y = \beta_{\text{year}} + f_1(\text{SBS}) + f_2(\text{Dep}) + f_3(\text{SBS}) \quad (1)$$

$$M2: Y = \beta_{\text{year}} + f_1(\text{SBS}) + f_2(\text{Dep}) + f_3(\text{SBS}) + f_5(\text{Lon}) + f_6(\text{Lat}) \quad (2)$$

$$M3: Y = \beta_{\text{year}} + f_1(\text{SBS}) + f_2(\text{Dep}) + f_3(\text{SBS}) + f_5(\text{Lon, Lat}) \quad (3)$$

$$M4: Y = \beta_{\text{year}} + f_1(\text{SBS}) + f_2(\text{Dep}) + f_3(\text{SBS}) + f_{\text{year}}(\text{Lon, Lat}) \quad (4)$$

$$M5: Y = \beta_{\text{year}} + f_1(\text{SBS}) + f_2(\text{Dep}) + f_3(\text{SBS}) + f_{\text{spat}}(\text{Lon, Lat}) \quad (5)$$

式中, M1 为包括底温、底盐、深度和年份的传统 GAM, M2 使用样条平滑经度和纬度效应, M1~2 为常规 GAM 模型, 与空间结构模型进行比较。GAM 的表达式:

$$Y = \beta + \sum_{i=1}^n s_i(x_i) + \varepsilon \quad (6)$$

式中,  $Y$  为资源量,  $\beta$  为截距项,  $s_i(x_i)$  为模型对第  $i$  变量拟合的平滑样条,  $\varepsilon$  为独立同分布的误差项。

M3 在 Model1 的基础上使用二维张量惩罚样条拟合空间趋势面<sup>[20-21]</sup>, M4 为不同年份拟合相互独立的趋势面, 属于变系数模型<sup>[22]</sup>, M5 应用地统计学的相关方法, 假设误差分布服从平稳高斯随机场 (stationary gaussian random fields), 服从均值为 0 的多元正态分布, 利用简化后的 Matérn 协方差函数对误差协方差矩阵进行参数估计, 属于空间加性模型<sup>[23]</sup>:

$$C(d_{ij}) = \sigma^2 \left( 1 + \frac{\sqrt{3}|d_{ij}|}{\rho} \right) e^{-\frac{\sqrt{3}|d_{ij}|}{\rho}} \quad (7)$$

式中,  $d_{ij}$  为站点  $i$  和站点  $j$  之间的距离,  $\sigma^2$  和  $\rho$  为待估计参数。

为防止过度拟合并保证各模型之间的可比性, 本研究将各模型的一维非参数平滑样条的节点参数  $K$  值限制为 5, 二维非参数平滑样条的  $K$  值限制为 15<sup>[3]</sup>。

调查数据显示, 海州湾小黄鱼资源量各年份的零值比例均在 20% 以上, 在 2014 年和 2015 年高达 50% 左右 (图 2)。本研究采用 delta 方法解决零值过多问题, 该方法也称 hurdle 方法或 two-step 方法, 被广泛应用与生态学和渔业科学<sup>[22,24-25]</sup>。delta 方法将观测渔获量  $Y_i$  分为两部分: 一是二项模型估计非零渔获率  $\pi_i$ , 二是正态模型估计非零渔获量  $\mu_i$ :

$$Z_i \sim \text{Bernoulli}(\pi_i) \quad (8)$$

$$\lambda_i \sim \text{Normal}(\mu_i) \quad (9)$$

$$\hat{Y}_i = Z_i \lambda_i \quad (10)$$

二项模型选用二项分布和 logit 连接函数预测渔获率  $Z_i$ , 正态模型可选用伽马分布, 正态分布或 tweedie 分布等多种分布族, 本研究经过比较选择正态分布和 identical 连接函数预测非零渔获量  $\lambda_i$ , 预测渔获量  $\hat{Y}_i$  为两部分的乘积。

### 1.3 模型效果评价

本研究根据受试者特征曲线下面积 (AUC)、

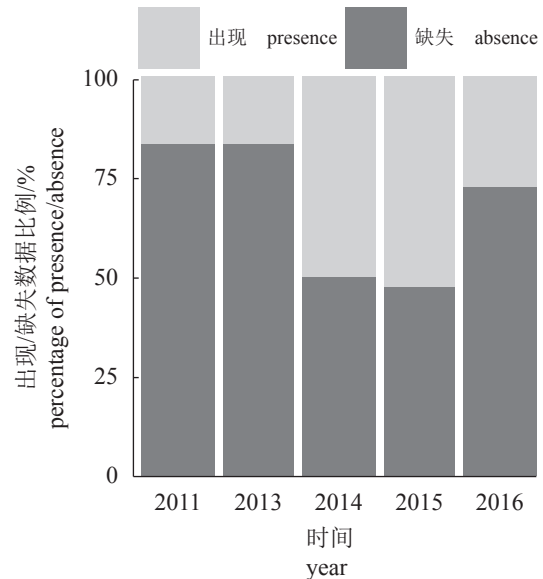


图 2 2011、2013—2016 年海州湾小黄鱼调查数据分布图

Fig. 2 Distribution of *L. polyactis* survey data in Haizhou Bay in 2011, 2013-2016

决定系数 ( $R^2$ ) 和赤池信息准则 (Akaike information criterion, AIC) 对不同模型的拟合效果进行对比。其中 AUC 和  $R^2$  量化模型的拟合效果; AIC 作为模型整体的评价指标, 其计算公式: <sup>[26]</sup>

$$\text{AIC} = 2k - 2\ln(\hat{L}) \quad (11)$$

式中,  $k$  为模型参数个数,  $\hat{L}$  模型似然函数。

残差空间自相关 (residual spatial autocorrelation) 表示邻近的点位具有相近的残差, 从而会影响建模过程和参数推断<sup>[27]</sup>。残差空间自相关性通过莫兰自相关图 (Moran's I correlograms) 进行可视化, 其反映了不同带宽下, 全局莫兰指数 (global Moran's index) 的变化情况, 带宽的选择通常为邻近点对数、滞后距离和空间位置关系等<sup>[28]</sup>, 本研究以不同邻近权重点对数构建全局莫兰指数, 全局莫兰指数计算公式:

$$\text{global Moran's index} = \frac{n}{s} \frac{\sum_{i=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (12)$$

$$s = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$$

式中,  $x_i$  为  $i$  站位的残差,  $\bar{x}$  为残差的算术平均值,  $n$  为站位数 (个),  $w$  为空间权重矩阵 (当两个站点定义为相邻时, 权重为 1, 反之为 0)。全局莫兰指数反映空间变量的聚集特性, 其值为正值, 代表空间相似; 负值代表空间相异; 计算结果趋于

零, 则代表空间随机<sup>[27]</sup>。

利用 300 次 5 折交叉验证 (5-fold cross validation) 检验所得模型的预测效果, 通过计算二项模型的 AUC 值, delta 模型的  $R^2$  以及均方根误差 (root mean squared error, RMSE) 等参数对模型针对两种数据类型 (渔获率和渔获量) 建模的稳健性、拟合和预测性能进行评价。

#### 1.4 分布预测

根据模型评估的结果, 选择表现最佳的模型和建模方法进行海州湾小黄鱼春季资源分布进行预测。用于预测的环境数据提取自无结构网格有限体积海洋海岸模型 (finite-volume coastal ocean model, FVCOM), FVCOM 是一种无结构网格、有限体积、自由表面三维原始方程的海洋环流模型<sup>[29]</sup>。本实验对调查海域进行网格化, 获得 1 321 个网格, 并对 FVCOM 数据进行网格平均化处理, 提取预测环境数据以预测海州湾小黄鱼春季资源分布并绘图。

本实验中数据处理、模型构建、绘图等工作在 R(version 3.6.2) 中进行, 利用  $R^2$ BayesX<sup>[30]</sup> 构建模型, 莫兰指数计算通过 `spdep`<sup>[31]</sup> 进行分析, 采用 `sp`<sup>[32]</sup> 进行网格化处理。

## 2 结果

### 2.1 模型比较

5 种模型的拟合效果显示, 加入空间因子的模型 (M2~M5) 在模型的拟合上具有明显优势, 且正态模型和 delta 模型在加入空间因子后其拟合效果的提升比二项模型更为明显。二项模型中, M4 (变系数模型) 拟合效果最佳, 其 AUC 值为 0.94, 较 M1 提高了 0.09, 在正态模型和 delta 模型中,  $R^2$  最高的模型分别为 M5 (空间加性模型) 和 M4, 较 M1 提高了约 0.3。加入空间因子的模型 AIC 值更低, 其中 M5 在二项模型和正态模型的 AIC 值均为最低值, 较 M1 分别下降了 7.60 和 144.90 (表 1)。

表 1 5 种模型拟合效果

Tab. 1 Fitting effect of 5 models

模型 models	加入的因子 added factors	二项模型 binomial model			正态模型 positive model		delta模型 delta model	
		AIC	$\Delta$ AIC	AUC	AIC	$\Delta$ AIC	$R^2$	$R^2$
M1	—	98.13		0.86	615.69		0.26	0.30
M2	$f(\text{Lon})+f(\text{Lat})$	94.69	-3.45	0.89	594.50	-21.19	0.43	0.53
M3	$f(\text{Lon}, \text{Lat})$	99.97	1.84	0.87	599.12	-16.57	0.49	0.55
M4	$f_{\text{year}}(\text{Lon}, \text{Lat})$	98.45	0.32	0.94	590.54	-25.15	0.51	0.68
M5	$f_{\text{year}}(\text{Lon}, \text{Lat})$	90.53	-7.60	0.90	470.79	-144.90	0.54	0.57

注: - 表示模型未加入解释变量或AIC未发生变化

note: - indicates that no explanatory variables are added to the model or AIC does not change

本研究根据对临近权重个数定义数量的不同做出模型残差莫兰相关图 (图 3), 5 种二项模型的残差空间自相关性差距较小, 而正态模型和 delta 模型的差距显著, 加入空间因子后的模型残差空间自相关性显著降低, 其中 M5 的表现最优, 各模型全局莫兰指数随着临近权重个数的增大而减少的趋势明显。

### 2.2 交叉验证

对二项模型进行 300 次 5 折交叉验证的结果显示, M4 (变系数模型) 在拟合上具有明显的优势, 但其预测性能明显差于其他模型。M5 (空间加性模型) 在拟合和预测性能上均具有较大优势, 且除 M5 外的其他加入空间项的模型预测性能均差

于 M1 (图 4)。

对 delta 模型进行 300 次 5 折交叉验证的结果显示, 各模型均表现出较好的拟合效果, 其中加入空间项的模型  $R^2$  均高于 0.55, 其中 M4 最高 (0.73) (表 2)。5 种模型的 RMSE 差距较小, 其预测性能排序: M5>M2>M3>M1>M4。总体来看, M5 具有最低的 RMSE (21.33)。

### 2.3 分布预测

研究表明空间加性模型 (M5) 的二项和正态模型均具有最低的 AIC, 分别为 90.53 和 470.79; 其 AUC 值和  $R^2$  分别为 0.90 和 0.54, 交叉验证的结果显示, 在预测小黄鱼出现概率和生物量时, 空间加性模型的准确度均最高。

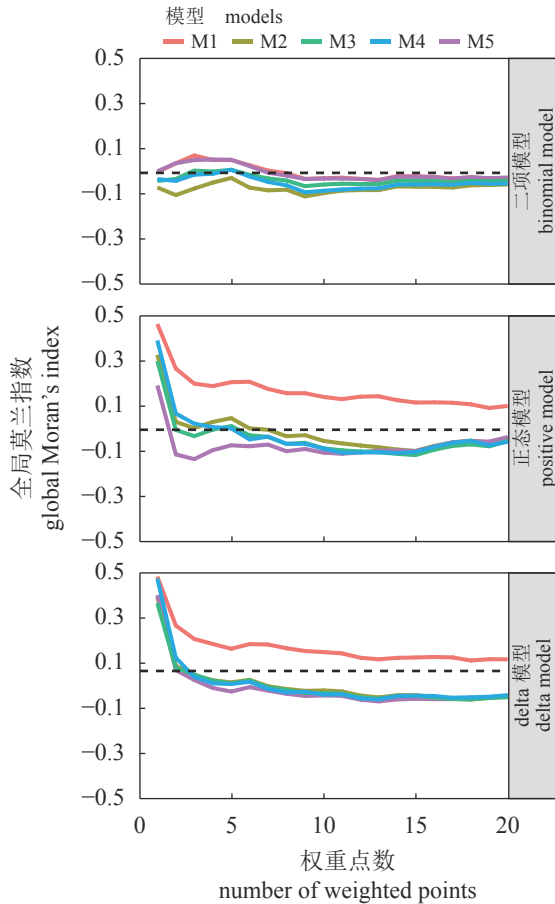


图3 模型残差莫兰相关图

Fig. 3 Moran correlograms of models residuals

基于 delta 空间加性模型，对海州湾小黄鱼春季相对资源分布进行预测 (图 5)。结果表明，其资源分布主要集中于南部和西部近岸地区，随着水深的增加而逐渐减少，且年间变动明显。此外，2011 年小黄鱼分布较为广泛，主要集中于 35°N 以南的海域。而在 2013—2016 年呈现资源量逐年递减的趋势并且其分布呈现出两个分布重心，分别位于西部近岸和南部海域，其中在西部近岸的分布较少且年间变化明显，南部海域的分布较广且年间变化较小。

### 3 讨论

#### 3.1 小黄鱼的时空分布

海州湾是黄海小黄鱼重要的产卵场之一，且春季是黄海小黄鱼的产卵期，因此，查明小黄鱼春季资源时空分布对海州湾小黄鱼的资源保护和持续利用至关重要。研究结果表明 delta 空间加性模型的预测效果最准确，预测结果显示，小黄鱼

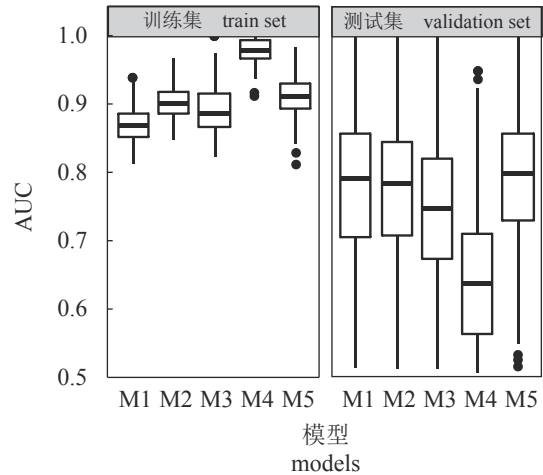


图4 binomial 模型交叉验证结果

Fig. 4 Cross-validation results of binomial models

表 2 delta 模型交叉验证结果

Tab. 2 Cross-validation results of delta models

模型 models	决定系数 $R^2$	均方根误差 RMSE
M1	0.31±0.05	22.12±5.61
M2	0.56±0.05	21.55±4.65
M3	0.60±0.07	21.65±4.83
M4	0.73±0.05	23.01±6.14
M5	0.58±0.08	21.33±5.33

的资源分布呈现出一定的时空格局和明显的年间变化。本研究中小黄鱼春季分布趋势的年间变化与邹易阳等<sup>[16]</sup>在海州湾小黄鱼栖息地适宜性研究，和仲霞铭等<sup>[33]</sup>对江苏近岸海域小黄鱼时空分布特征的研究中得出的结论一致。小黄鱼主要分布在 35°N 以南、10~20 m 等深线处的江苏沿岸海域，并在此进行产卵、孵化，这种分布格局与海州湾理化环境与地理环境特点密切相关。相关研究表明，小黄鱼的产卵区一般分布在河口区和入海径流较大的沿海区，且每年 4—5 月，随水温回升，在外海越冬的小黄鱼进入近岸渔场产卵，适宜其产卵的底层海水温度为 16~22 °C，底层海水盐度是 31.59~34.65<sup>[19, 34-35]</sup>。受黄海冷水团的影响，海州湾春、夏两季东北部海域底层水温明显偏低，而小黄鱼产卵群体主要分布在水团交界处，因为温度、盐度梯度可能会刺激小黄鱼性腺发育<sup>[36]</sup>，此外，海州湾南部近岸海域够为小黄鱼的春季产卵和幼体的生长和发育提供合适的泥沙底质、高温、低盐等环境条件，同时 35°N 以南的近岸海域

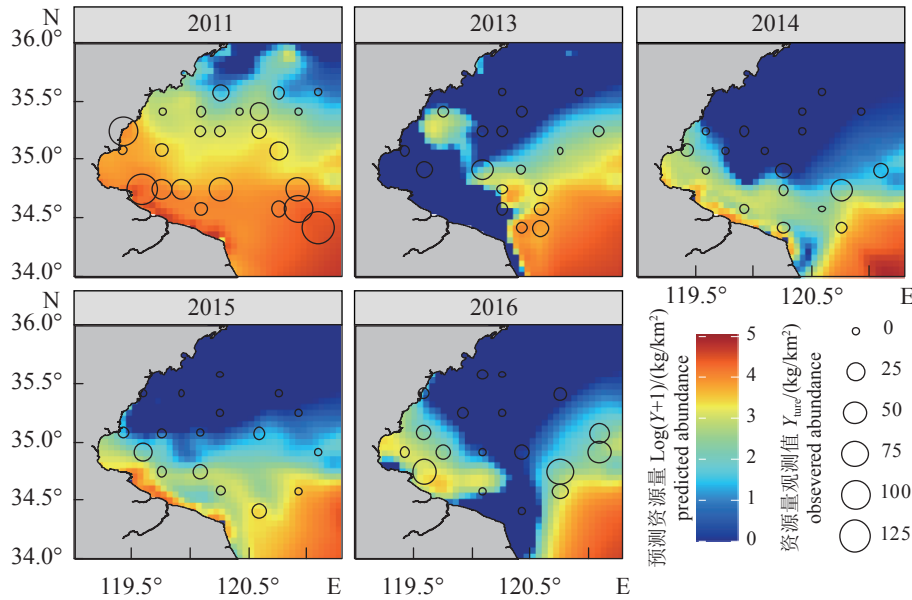


图 5 delta 空间加性模型预测的海州湾小黄鱼春季资源分布

Fig. 5 Predicted abundance distribution of *L. polyactis* in Haizhou Bay during spring by delta-geoaddivitive model

为淮河、射阳河等多条河流入海口, 受到陆地径流和沿岸流的影响, 水中丰富的营养盐能够为小黄鱼的产卵繁殖提供适宜的营养条件。产卵后的小黄鱼群体摄食需求强烈, 同时幼鱼也在产卵场附近进行索饵<sup>[15]</sup>, 这可能是小黄鱼在海州湾南部近岸海域的分布呈现出多个热点的原因之一。

### 3.2 STAR 模型

本研究通过 STAR 模型框架分析了空间因子对 GAM 拟合以及预测性能的影响。研究表明, 空间因子的加入能够优化模型拟合效果, 5 种模型的预测性能差距较小, 加入空间因子的模型表现出低残差空间自相关性, 增强了模型的残差独立性并使其参数估计更加可靠, 因此空间因子的加入对回归模型的构建起正向作用。其中, 空间加性模型利用地统计方法对误差项进行拟合, 拥有较高拟合效果的同时保持了模型的复杂程度和残差空间自相关性在较低的水平, 而变系数模型则出现了过度拟合的现象, 其拟合效果最优但其预测性能最差。Potts 等<sup>[3]</sup>比较了几种不同空间效应输入方式对 GAM 的影响, 发现趋势面 GAM 能够在拟合效果和预测性能之间达到较好的权衡。Stock 等<sup>[22]</sup>比较了趋势面 GAM 和变系数模型在预测夏威夷延绳钓渔业副渔获率的表现, 其结论和本实验一致, 变系数模型出现了过度拟合的现象, 在多数物种的预测表现较趋势面 GAM 更差。尽

管如此, 变系数模型能够解释变量之间交互作用的特点在渔业中也有一定的应用, 如 Dippold 等<sup>[37]</sup>通过变系数模型对单位捕捞努力渔获量 (catch per unit effort, CPUE) 建模发现在不同的空间分区下, 目标变量对种群动力学参数的响应不同。基于地统计学的物种分布模型通过空间随机效应解释了未被观测到的决定物种空间分布的过程, 以往的研究表明其在适宜地理尺度上拥有较好的预测性能<sup>[6,38-39]</sup>, 与本研究结论一致。

本研究应用 delta 方法处理调查数据中零值过多的问题, 结果显示其效果较好。但同时应该考虑该方法的适用性: 一是样本量大小和零值的比率, 二是 delta 方法假设存在两个相互独立的过程决定渔获率和渔获量, 而在 Thorson<sup>[40]</sup>和 Cantoni 等<sup>[41]</sup>的研究中都表明其假设可能不成立。在未来的工作中, 可以尝试其他处理零膨胀问题的方法, 如尝试 Thorson 改进后的 delta 模型<sup>[40]</sup>, 或允许偏态分布的指数族分布如 gamma 分布、tweedie 分布等。

通过物种分布模型来探究物种分布和相关变量之间的关系, 从而预测目标物种的资源分布是实现该目的的有效手段之一。本研究考虑并对比了几种 STAR 模型, 结合常见的零值处理方法 delta 方法, 对海州湾小黄鱼春季资源时空分布进行了建模分析。但本研究考虑的因子较少, 如饵料生物、海洋环流、底质类型等因子未能加入其中,

除此之外本研究仅考虑了空间自相关对建模的影响,未能考虑时间自相关,今后可以考虑通过加入自回归过程、随机效应等方式加入时间效应以提高模型的精确程度<sup>[42]</sup>。

(作者声明本文无实际或潜在的利益冲突)

### 参考文献 (References):

- [1] Naimi B, Araújo M B. sdm: a reproducible and extensible R platform for species distribution modelling[J]. *Ecography*, 2016, 39(4): 368-375.
- [2] Robinson N M, Nelson W A, Costello M J, et al. A systematic review of marine-based species distribution models (SDMs) with recommendations for best practice[J]. *Frontiers in Marine Science*, 2017, 4: 421.
- [3] Potts S E, Rose K A. Evaluation of GLM and GAM for estimating population indices from fishery independent surveys[J]. *Fisheries Research*, 2018, 208: 167-178.
- [4] 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.
- [5] Solanki H U, Bhatpuria D, Chauhan P. Applications of generalized additive model (GAM) to satellite-derived variables and fishery data for prediction of fishery resources distributions in the Arabian Sea[J]. *Geocarto International*, 2017, 32(1): 30-43.
- [6] Maunder M N, Thorson J T, Xu H K, et al. The need for spatio-temporal modeling to determine catch-per-unit effort based indices of abundance and associated composition data for inclusion in stock assessment models[J]. *Fisheries Research*, 2020, 229: 105594.
- [7] Brezger A, Lang S. Generalized structured additive regression based on Bayesian P-splines[J]. *Computational Statistics & Data Analysis*, 2006, 50(4): 967-991.
- [8] Maunder M N, Punt A E. Standardizing catch and effort data: a review of recent approaches[J]. *Fisheries Research*, 2004, 70(2-3): 141-159.
- [9] Lo N C H, Jacobson L D, Squire J L. Indices of relative abundance from fish spotter data based on delta-lognormal models[J]. *Canadian Journal of Fisheries and Aquatic Sciences*, 1992, 49(12): 2515-2526.
- [10] 孟庆闻, 缪学祖, 俞泰济, 等. 鱼类学 [M]. 上海: 上海科学技术出版社, 1989: 1-298.
- [11] Meng Q W, Miao X Z, Yu T J, et al. *Ichthyology*[M]. Shanghai: Shanghai Scientific & Technical Publishers, 1989: 1-298 (in Chinese).
- [12] 郑元甲, 陈雪忠, 程家骅, 等. 东海大陆架生物资源与环境 [M]. 上海: 上海科学技术出版社, 2003: 1-834.
- [13] Zheng Y J, Chen X Z, Cheng J H, et al. The biological resources and environment in continental shelf of the East China Sea[M]. Shanghai: Shanghai Scientific and Technic Publishers, 2003: 1-834(in Chinese).
- [14] 林龙山, 程家骅. 延长东海区伏季休渔期的渔业效果分析[J]. 大连水产学院学报, 2009, 24(1): 12-16.
- [15] Lin L S, Cheng J H. Effects of the prolonged summer closed fishing period on fisheries in East China Sea[J]. *Journal of Dalian Fisheries University*, 2009, 24(1): 12-16 (in Chinese).
- [16] 刘尊雷, 杨林林, 袁兴伟, 等. 基于集成模型的小黄鱼越冬群体适宜生境及其环境影响因素[J]. 应用生态学报, 2020, 31(6): 2076-2086.
- [17] Liu Z L, Yang L L, Yuan X W, et al. Overwintering distribution and its environmental determinants of small yellow croaker based on ensemble habitat suitability modeling[J]. *Chinese Journal of Applied Ecology*, 2020, 31(6): 2076-2086 (in Chinese).
- [18] 刘梦影, 吕泽华, 田思泉, 等. 应用栖息地适宜性指数研究温台渔场小黄鱼的空间分布特征[J]. *上海海洋大学学报*, 2020, 29(1): 92-101.
- [19] Liu M Y, Lv Z H, Tian S Q, et al. Spatial distribution of *Larimichthys polyactis* in Wen-Tai fishing ground based on habitat suitability index[J]. *Journal of Shanghai Ocean University*, 2020, 29(1): 92-101 (in Chinese).
- [20] 尹洁, 王晶, 张崇良, 等. 利用two-stage GAM研究海州湾及其邻近海域小黄鱼鱼卵的时空分布特征[J]. *中国水产科学*, 2019, 26(6): 1164-1174.
- [21] Yin J, Wang J, Zhang C L, et al. Spatial and temporal distribution characteristics of *Larimichthys polyactis* eggs in Haizhou Bay and adjacent regions based on twostage GAM[J]. *Journal of Fishery Sciences of China*, 2019, 26(6): 1164-1174 (in Chinese).
- [22] 邹易阳, 薛莹, 麻秋云, 等. 应用栖息地适宜性指数研究海州湾小黄鱼的空间分布特征[J]. *中国海洋大学学报(自然科学版)*, 2016, 46(8): 54-63.
- [23] Zou Y Y, Xue Y, Ma Q Y, et al. Spatial distribution of *Larimichthys polyactis* in Haizhou Bay based on habitat suitability index[J]. *Periodical of Ocean University of China* 中国水产学会主办 sponsored by China Society of Fisheries

- China, 2016, 46(8): 54-63 (in Chinese).
- [17] Jordaan A, Chen Y, Townsend D W, *et al.* Identification of ecological structure and species relationships along an oceanographic gradient in the Gulf of Maine using multivariate analysis with bootstrapping[J]. *Canadian Journal of Fisheries and Aquatic Sciences*, 2010, 67(4): 701-719.
- [18] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. GB/T 12763.6—2007 海洋调查规范 第 6 部分: 海洋生物调查 [S]. 北京: 中国标准出版社, 2008: 6-17.
- General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of China. GB/T 12763.6-2007 Specifications for oceanographic survey-Part 6: marine biological survey[S]. Beijing: China Standards Press, 2008: 6-17 (in Chinese).
- [19] 张云雷, 徐宾铎, 张崇良, 等. 基于 Tweedie-GAM 模型研究海州湾小黄鱼资源丰度与栖息环境的关系[J]. *海洋学报*, 2019, 41(12): 78-89.
- Zhang Y L, Xu B D, Zhang C L, *et al.* Relationship between the habitat factors and the abundance of small yellow croaker (*Larimichthys polyactis*) in Haizhou Bay based on the Tweedie-GAM model[J]. *Haiyang Xuebao*, 2019, 41(12): 78-89 (in Chinese).
- [20] Jones K, Wrigley N. Generalized additive models, graphical diagnostics, and logistic regression[J]. *Geographical Analysis*, 1995, 27(1): 1-18.
- [21] Dormann C F, McPherson J M, Araújo M B, *et al.* Methods to account for spatial autocorrelation in the analysis of species distributional data: a review[J]. *Ecography*, 2007, 30(5): 609-628.
- [22] Stock B C, Ward E J, Eguchi T, *et al.* Comparing predictions of fisheries bycatch using multiple spatiotemporal species distribution model frameworks[J]. *Canadian Journal of Fisheries and Aquatic Sciences*, 2019, 77(1): 146-163.
- [23] Kammann E E, Wand M P. Geoadditive models[J]. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 2003, 52(1): 1-18.
- [24] Arcuti S, Pollice A, Ribocco N, *et al.* Bayesian spatiotemporal analysis of zero-inflated biological population density data by a delta-normal spatiotemporal additive model[J]. *Biometrical Journal*, 2016, 58(2): 372-386.
- [25] Lynch H J, Thorson J T, Shelton A O. Dealing with under- and over-dispersed count data in life history, spatial, and community ecology[J]. *Ecology*, 2014, 95(11): 3173-3180.
- [26] Akaike H. Information theory and an extension of the maximum likelihood principle[M]//Parzen E, Tanabe K, Kitagawa G. Selected Papers of Hirotugu Akaike. New York: Springer, 1998: 199-213.
- [27] Gaspard G, Kim D, Chun Y. Residual spatial autocorrelation in macroecological and biogeographical modeling: a review[J]. *Journal of Ecology and Environment*, 2019, 43(1): 19.
- [28] Mathur M. Spatial autocorrelation analysis in plant population: an overview[J]. *Journal of Applied and Natural Science*, 2015, 7(1): 501-513.
- [29] 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.
- [30] Umlauf N, Adler D, Kneib T, *et al.* Structured additive regression models: an R interface to BayesX[J]. *Journal of Statistical Software*, 2015, 63(21): 1-46.
- [31] Bivand R S, Wong D W S. Comparing implementations of global and local indicators of spatial association[J]. *TEST*, 2018, 27(3): 716-748.
- [32] Pebesma E, Bivand R S. S. Classes and methods for spatial data: the sp package[CP/OL]. <https://cran.r-project.org/doc/news/>.
- [33] 仲霞铭, 张虎, 汤建华, 等. 江苏近岸海域小黄鱼时空分布特征[J]. *水产学报*, 2011, 35(2): 238-246.
- Zhong X M, Zhang H, Tang J H, *et al.* Temporal and spatial distribution of *Larimichthys polyactis* Bleeker resources in offshore areas of Jiangsu Province[J]. *Journal of Fisheries of China*, 2011, 35(2): 238-246 (in Chinese).
- [34] 金显仕, 赵宪勇, 孟田湘, 等. 黄、渤海生物资源与栖息环境 [M]. 北京: 科学出版社, 2005: 1-405.
- Jin X S, Zhao X Y, Meng T X, *et al.* Yellow Sea and Bohai Sea biological resources and habitats[M]. Beijing: Science Press, 2005: 1-405 (in Chinese).
- [35] 陈佳杰, 徐兆礼, 陈雪忠. 我国沿海小黄鱼渔场的空间格局[J]. *水产学报*, 2010, 34(2): 236-244.
- Chen J J, Xu Z L, Chen X Z. The spatial distribution pattern of fishing ground for small yellow croaker in China Seas[J]. *Journal of Fisheries of China*, 2010, 34(2): 236-



- 244 (in Chinese).
- [36] 丁峰元, 林龙山, 李建生, 等. 东海区北部小黄鱼生殖群体分布及与水团关系[J]. 自然资源学报, 2007, 22(6): 1013-1019.
- Ding F Y, Lin L S, Li J S, *et al.* Relationship between redlip croaker (*Larimichthys polyactis*) spawning stock distribution and water masses dynamics in Northern East China Sea Region[J]. Journal of Natural Resources, 2007, 22(6): 1013-1019 (in Chinese).
- [37] Dippold D A, Adams G D, Ludsins S A. Spatial patterning of walleye recreational harvest in Lake Erie: role of demographic and environmental factors[J]. *Fisheries Research*, 2020, 230: 105676.
- [38] Thorson J T, Adams G, Holsman K. Spatio-temporal models of intermediate complexity for ecosystem assessments: a new tool for spatial fisheries management[J]. *Fish and Fisheries*, 2019, 20(6): 1083-1099.
- [39] Thorson J T, Skaug H J, Kristensen K, *et al.* The importance of spatial models for estimating the strength of density dependence[J]. *Ecology*, 2015, 96(5): 1202-1212.
- [40] Thorson J T. Three problems with the conventional delta-model for biomass sampling data, and a computationally efficient alternative[J]. *Canadian Journal of Fisheries and Aquatic Sciences*, 2018, 75(9): 1369-1382.
- [41] Cantoni E, Flemming J M, Welsh A H. A random-effects hurdle model for predicting bycatch of endangered marine species[J]. *The Annals of Applied Statistics*, 2017, 11(4): 2178-2199.
- [42] Cavieres J, Nicolis O. Using a spatio-temporal Bayesian approach to estimate the relative abundance index of yellow squat lobster (*Cervimunida johni*) off Chile[J]. *Fisheries Research*, 2018, 208: 97-104.

## Spatio-temporal distribution of *Larimichthys polyactis* in Haizhou Bay based on STAR model

ZHAO Wei<sup>1,2</sup>, REN Yiping<sup>1,2,3</sup>, XU Binduo<sup>1,2</sup>, XUE Ying<sup>1,2</sup>, ZHANG Chongliang<sup>1,2\*</sup>

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

2. Field Observation and Research Station of Haizhou Bay Fishery Ecosystem, Ministry of Education, Qingdao 266003, China;

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

**Abstract:** This study was conducted to evaluate and improve model performance for estimating abundance and occurrence of small yellow croaker (*Larimichthys polyactis*) based on the data collected from fishery independent bottom trawl surveys data in Haizhou Bay in the spring of 2011 and 2013-2016. According to different methods of processing spatial data, five species distribution models were formulated by combining the delta method with structured additive regression framework, they were compared in terms of their performance on fitness, predictive capacity and independence of residuals for two commonly used response variables, namely, occurrence and abundance. Result showed that models with spatial covariates had significantly better fitting effect and lower residual spatial correlation, and positive model and delta model showed more improvement than binomial model. The AIC of geoadditive models was the lowest in both binomial model and positive model (respectively lower than GAM without the spatial term by 7.60 and 144.90). Varying coefficient models had the highest  $R^2$  (0.68) and fitting AUC (0.94). Geoadditive model had the highest AUC (0.793±0.100) and the lowest RMSE (21.65±4.83), indicating that geoadditive model's estimation on the occurrence probability and resource density of *L. polyactis* were most accurate. Therefore, we predicted the spatial distribution of *L. polyactis* in Spring using delta-geoadditive model based on FVCOM simulation data. Prediction result reflected the distribution and variations of *L. polyactis* was mainly distributes in the southern and western coastal areas (34.0°N-34.5°N, 121.0°E-121.5°E), decreasing with the water depth increase and varying significantly over the years. This study aimed to provided a scientific basis for the development and protection of *L. polyactis* fishery resources in Haizhou Bay.

**Key words:** *Larimichthys polyactis*; structured additive regression; spatial effect; cross-validation; model comparison; Haizhou Bay

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

**Funding projects:** Marine S & T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao)(2018SDKJ0501-2); National Key R & D Program of China (2018YFD0900904); National Natural Science Foundation of China (31772852)