文章编号：1674-5566(2024)06-1357-12 DOI：10. 12024/jsou. 20231004323

基于机器学习的西南印度洋深海散射层声学资源密度预测

万树杰[1]，陈新军[1,2,3,4]

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

摘 要：预测大洋中深海散射层的资源丰度与分布，对指示海洋保护动物与重要渔场分布，开发散射层中的渔业资源等具有重要意义。本研究以海里面积散射系数(Nautical area scattering coefficient，NASC)作为散射层的资源密度指标，综合多个环境因子，利用K-means聚类和SSA-XGBoost模型，实现了对西南印度洋散射层资源密度的分类预测。结果表明，模型预测的准确率为80. 51%，精确率为76%，召回率为78%，样本数据与预测数据的高密度区域空间分布相匹配，模型的应用效果较好。通过对2011年不同季节散射层密度的预测，发现散射层高密度区的重心由东南向西北方向移动，其中春季时期高密度区的重心的纬度最大，冬季时高密度区的重心的纬度最小，高密度区的重心点在东南-西北方向上的离散性大于东北-西南方向。本研究可为阐明散射层大尺度空间分布和资源变动规律提供新的方法。

关键词：深海散射层；机器学习；声学资源密度；西南印度洋

中图分类号：S 932. 4 **文献标志码：**A

深海散射层(Deep scattering layer，DSL；简称散射层)自1942年在太平洋被发现以来，已经被证明在世界的各个大洋中广泛存在[1-3]。散射层在海洋食物网中扮演着至关重要的角色，其中蕴藏着的丰富的海洋生物资源是海洋生物群落中最大的群体之一，主要由不同规格的浮生动物和中层鱼(Mesopelagic fishes)等聚集形成，包括小型鱼类、甲壳类、头足类和胶状生物，体长从2 cm到20 cm不等[4-5]。这些生物一般夜间由中层(200~1 000 m)游到上层(0~200 m)觅食，日出后再下潜回到中层以躲避大型捕食者[6-7]。由于散射层生物体型较小，在拖网采样调查中容易逃脱，通常基于拖网调查的海洋中上层散射层的生物量往往被低估，因此它们在海洋生物地球化学循环和食物网中的作用可能远远大于目前的认知[8]。国内有关散射层的研究可追溯到李玉昕等[9]在南海进行的关于深海散射层的实验研究，研究表明南海海域的散射层是由尺寸为几厘米的有鳃鱼群体构成。此后，裘辛方等[10]采用垂直探鱼仪对黄海中部夏季反向体积声散射进行了观察，发现黄海中部海域晚上散射很强，白天则很弱。刘世刚等[11]研究发现南海深海存在2个明显的声学散射层，并且2个散射层间存在较为明显的昼夜垂直移动现象。陈钊等[12]对南海秋季声散射层的垂向分布特征和日变化进行了分析，结果表明，南海存在着2个声散射层，并估算出的散射层中散射体的垂向迁移速度。张超等[13]通过对西太平洋声散射层的垂向分布特征和日变化规律开展研究，发现西太平洋存在着2个声散射层，2个声散射层散射强度具有明显的日变化特征。高爽等[14]对南海北部声散射的季节变化研究发现，南海北部水体在垂向上分布着上散射层和深散射层2个主要散射层，且2个散射层的距离在夏季最远，在春秋最近。

声学调查方法由于具有调查时间短、覆盖范围大和不干扰调查对象的原始状态等特点，已被广泛应用于散射层资源丰度和生态特征等的研究中[15-18]。尽管声学方法调查散射层资源丰度具

收稿日期：2023-10-07 **修回日期：**2024-01-05

基金项目：国家重点研发计划(2019YFD0901401)

作者简介：万树杰(1994—)，男，博士研究生，研究方向为渔业声学。E-mail：shujie.wan@foxmail.com

通信作者：陈新军，E-mail：xjchen@shou.edu.cn

有快速、高效等优点，但由于散射层空间分布广泛，考虑到声学调查的经济性，大规模的散射层声学调查也往往难以开展，因此调查一般不能覆盖整个海域，通常是集中在某一区域。研究表明[19]，散射层分布由于受到各种海洋环境因素的影响，从而导致其资源丰度分布的空间和时间存在差异，调查结果通常不能代表整个海域散射层的资源分布状况。因此，基于现有调查数据预测不同海域、不同时间散射层的资源丰度，对深入了解更大空间和时间尺度上散射层的资源变动规律具有一定的现实意义。

随着大数据时代的到来，机器学习在许多领域得到了广泛的应用。如在海洋科学及其交叉领域的研究中，机器学习已经被应用于海水温度预报[20]、海水深度反演[21]、赤潮发生预警[22]、红树林生物量反演[23]、风暴潮预报等研究[24]中，均取得了良好的应用效果。而在渔业方面的研究中，准确的渔情渔场预报结果可以指导渔船的捕捞作业，减少盲目找鱼的时间，创造经济和生态效益。如魏广恩等[25]采用随机森林对北太平洋柔鱼资源丰度进行预测；常亮等[26]基于BP神经网络构建西北太平洋柔鱼资源丰度预测模型；周茜涵等[27]基于优化灰色模型对南海鸢乌贼资源丰度进行预测。而深度学习方法在散射层资源丰度预测的相关研究相对较少，为了探索散射层的资源丰度预测方法，本研究以2011—2020年间西南印度洋的声学调查数据为基础，综合多个环境因子，建立了基于机器学习的散射层资源密度预测模型。基于该模型，进一步预测了2011年不同季节的散射层的资源密度分布情况，并分析了其重心变动规律。

1 材料与方法

1.1 数据来源

1.1.1 声学数据

本研究所采用的声学数据集来自澳大利亚海洋综合观测系统（Australia′s integrated marine observing system，IMOS）中的生物声学观测计划[28]。该观测系统开始于2010年，主要采集不同商业渔船及科考船上装备的SIMARD系列鱼探仪的声学数据，包括ES60、ES70型号，相关鱼探仪均完成了科学校准工作。

如图1所示，研究选取2011—2020年间西南印度洋109条声学调查断面数据，调查海域为13°S~54°S和44°E~111°E。研究选取了该数据集中38 kHz的声学数据，数据处理流程基于RYAN等[29]制定的声学数据处理框架，主要通过滤波的方式剔除海洋环境中的间歇性噪声尖峰、持续的间歇性噪声、衰减脉冲和最终的背景噪声。积分单元按照1 000 m为间隔对10~1 000 m水层进行积分。通过对数据的分析处理，最终的数据输出单元为水平距离（1 000 m）×水深（10~1 000 m，垂向每10 m为一个数据输出单元）的水体体积后向散射强度（Mean volume backscattering strength，S_v），阈值设定为−130 dB。该数据集直接提供声学处理后的结果即S_v值。该值是对采样水体内生物个体的后向散射强度的线性相加，因此其与水体内主要散射生物的资源密度成正比，其计算公式为

$$S_v[i,j] = P_{er}[i,j] + \lg r[i,j] - 2\alpha_a r[i,j] - 10\lg\left(\frac{(P_{et}\lambda^2 g_0{}^2 c_w \tau\psi)}{32\pi^2}\right) - 2S_{acorr} \quad (1)$$

式中：i和j分别为垂直采样数目和水平脉冲数目；P_{er}为在距离r处能够接收到调查目标的最小回声强度，dB；r为到探测目标的距离，m；α_a为声波衰减系数，dB/m；P_{et}为换能器功率，W；λ为波长，m；g_0为声学校正的换能器增益，dB；c_w为海水中的声速值，m/s；τ为脉冲宽度，s；ψ为等效双向波束角，dB；S_{acorr}为声学校正获得的回声积分值修正参数，dB。

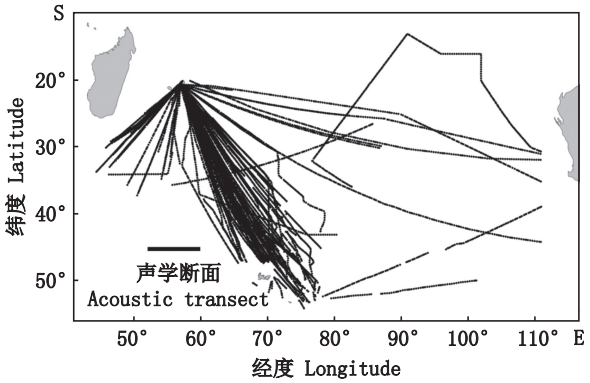

图1 声学断面空间分布

Fig. 1 Spatial distribution of acoustic transects

在得到S_v的基础上，进一步计算相应的海里面积散射系数（Nautical area scattering coefficient，NASC），作为本研究中声散射层的资源丰度值，其计算公式[30]：

$$S_A = 4\pi(1\,852)^2 \times 10^{S_v/10} \times T \quad (2)$$

式中：S_A为海里面积散射系数，m^2/nmi^2；S_v为水体体积后向散射强度；T为积分栅格的厚度。

1.1.2　环境数据

研究共选取了8个环境因子数据（表1），包括叶绿素a质量浓度（Chlorophyl-a，*Chl.*a）、溶解氧（Dissolved oxygen，DO）、pH、混合层深度（Mixed layer depth，MLD）、海表面温度（Sea surface temperature，SST）、光和有效辐射（Photosynthetically active radiation，PAR）、涡动能（Eddy kinetic energy，EKE）和水深（Depth）。环境因子与NASC值的空间匹配，基于ArcGIS软件完成。

表1　环境因子

Tab. 1　Environmental variables used in this study

环境因子 Variables	空间分辨率 Spatial resolution	时间分辨率 Temporal resolution	单位 Unit	数据来源 Source
叶绿素a *Chl.*a	0.25°×0.25°	月平均	mg/m^3	https://data.marine.copernicus.eu/
溶解氧 DO	0.25°×0.25°	月平均	$mmol/m^3$	
pH	0.25°×0.25°	月平均	无	
混合层深度 MLD	0.083°×0.083°	月平均	m	
海表面温度 SST	0.05°×0.05°	月平均	℃	
涡动能 EKE	0.083°×0.083°	月平均	m^2/s^2	
光合有效辐射 PAR	0.042°×0.042°	月平均	W/m^2	https://oceanwatch.pifsc.noaa.gov/
水深 Depth	0.033°×0.033°	无	m	https://rda.ucar.edu

1.1.3　数据重采样

声学数据集中，数据水平间隔为1 km，垂向间隔为10 m，数据分布密集。为方便数据的统计与分析，对数据集进行重采样。首先第一次重采样在垂直方向上，每一个采样点从20 m水深开始，每2组数据取1组，再将同一经纬度的垂直数据相加，作为该采样位置的NASC值，后在水平方向上每10组取1组。基于第一次重采样的结果，将研究区域在经纬度按20′为间隔进行划分，形成20′×20′的网格，通过统计每个网格内NASC值的个数与大小，计算NASC值的和作为网格的数值，网格中心点的经纬度为网格坐标点，对数据进行重采样。数据重采样主要基于ArcGIS软件完成。

1.2　实验设计

基于声学数据处理和环境数据的空间匹配及数据重采样结果，得到表征散射层资源丰度的指标NASC数据集，且每一个NASC值都有相对应的经纬度和环境因子（表1）。该数据集中，75%作为模型的训练样本，25%作为测试样本评估模型。

首先将样本进行标准化处理，再采用K-means聚类将标准化的样本进行分类，参考PROUD等[31]的研究，将样本共分为10类。以每个样本的经度、纬度、叶绿素a质量浓度、溶解氧、pH、混合层深度、海表面温度、光和有效辐射、涡动能和水深等10个变量作为输入数据，每个样本的类别为输出数据。基于训练样本的K-means聚类结果，分别计算每一类数据的NASC均值，作为后续预测结果中相应类别的NASC值。

基于训练好的模型，进一步对2011年西南印度洋海域（10°S~56°S，41°E~117°E）内不同季节的散射层声学资源密度进行预测，采用克里金插值方法对预测结果进行插值分析，绘制研究海域散射层资源密度的水平分布图，并采用标准差椭圆（方向分布）方法，分析散射层密度重心的空间分布趋势与变动规律。其中，标准差椭圆的长半轴和短半轴分别表示数据分布的方向和范围[32]。长短半轴的值差距越大，数据的方向性越明显。短半轴越短，数据呈现的向心力越明显，否则表示数据的离散程度越大[33]。

1.3　预测模型构建

1.3.1　极端梯度提升模型（eXtreme gradient boosting model，XGBoost）

XGBoost是机器学习中的一种集成学习算法，该模型在处理大数据问题中表现优秀[34]。XGBoost作为一种决策树算法，通过逐步训练多

个弱分类器来构建一个强大的集成分类器，该算法的主要思想是优化损失函数，以找到最佳模型[34]。XGBoost寻找最优参数组合的传统方法是首先根据经验为每个参数设置几个值，然后通过计算每个组合的准确率来选择准确率最高的组合，然后在一定的组合范围内，进行网格搜索，选择最优组合。由于模型参数较多，因此其参数寻优耗时较长。

XGBoost的数学模型如下：

假设$D_1=\{(x_i,y_i)\}$是由n个样本和m个特征值组成的数据集。附加函数z被集合树模型用来近似系统响应，如下：

$$\hat{y}_i=\varphi(x_i)=\sum_{z=1}^{Z}f_z(x_i),\ f_z\in F \tag{3}$$

式中：F为包含Z棵树的函数空间，被定义为

$$F=\{f(x)=\omega_{q(x)}\}(q:R^m\rightarrow T,\omega\in R^T) \tag{4}$$

式中：q为树的结构；T为叶子个数；ω为叶子的权重；$\omega_{q(x)}$为叶子节点q的分数；$f(x)$为某一独立树；f_z是为与q、ω相联系并于独立树相关的函数。

为了优化集合树预测性能，定义XGBoost的目标函数为

$$\begin{aligned}\partial(\theta)=L(y,\hat{y})+\Omega(\theta)&=\frac{1}{2}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\gamma T+\\&\frac{1}{2}\lambda\sum_{j=1}^{T}\omega_j^2=\frac{1}{2}\sum_{i=1}^{n}(y_i-\sum_{k=1}^{T}f_k(x_i)_i)^2+\gamma T+\\&\frac{1}{2}\lambda\sum_{j=1}^{T}\omega_j^2\end{aligned} \tag{5}$$

式中：L为显示预测误差的凸形损失函数；y_i是为真实值；k为误差最小化过程的迭代次数。

1.3.2 麻雀搜索算法（Sparrow search algorithm，SSA）

SSA是一种模拟麻雀群体觅食过程的算法，其通过模拟麻雀觅食及躲避捕食者的行为来优化模型参数，寻优能力强且收敛速度快。该算法可以有效减少搜索空间，提高搜索效率，避免搜索空间中的局部最优解[35-36]。

SSA的数学模型为将70%的麻雀分配为生产者，在每次迭代循环中，发现者的位置更新规则为

$$X_{i,j}^{t+1}=\begin{cases}X_{i,j}^{t}\times\exp\left(\dfrac{-i}{\alpha\times t_{\max}}\right),R_2<\mathrm{ST}\\X_{i,j}^{t}+Q\times L\quad,R_2\geqslant\mathrm{ST}\end{cases} \tag{6}$$

式中：t为当前迭代次数；$X_{i,j}$为第i只麻雀在第j维的位置；α为(0,1]内的随机数；$t_{\max}$为最大迭代次数；安全值ST为区间[0.5,1]中的随机数。预警值R_2为[0,1]中的随机数；随机数Q服从高斯分布；L为一个1×d的矩阵，d为优化变量的维度。

位置更新公式为

$$X_{i,j}^{t+1}=\begin{cases}X_b^{t+1}+|X_{i,j}^{t}-X_b^{t+1}|\times A^{+}\times L\ ,i\leqslant l/2\\Q\times\exp\left(\dfrac{X_w^{t}-X_{ij}^{t}}{i^2}\right),i>l/2\end{cases} \tag{7}$$

式中：X_b和X_w分别为发现者搜索的全局最优位置和全局最差位置；A为大小为1×d的矩阵，其中每个元素随机设置为-1或1，且$A^{+}=A^{\mathrm{T}}(AA^{\mathrm{T}})-1$，$d$为优化变量的维度。当$i\leqslant l/2$时，加入者与找到最佳位置的生产者竞争食物，$l$为麻雀的数量。否则，适应度较低的加入者处于饥饿状态，需要寻找新的位置觅食。

当意识到捕食者的危险时，麻雀的位置改变规则如下：

$$X_{i,j}^{t+1}=\begin{cases}X_{\mathrm{best}}^{t}+\beta\times|X_{i,j}^{t}-X_{\mathrm{best}}^{t}|,f_i>f_g\\X_{i,j}^{t}+K\times\left(\dfrac{|X_{ij}^{t}-X_{\mathrm{worst}}^{t}|}{(f_i-f_w)+\psi}\right),f_i=f_g\end{cases} \tag{8}$$

式中：X_{best}为当前全局最优位置；K为[-1,1]中的随机数；β为服从正态分布的步长控制参数；f_i、f_g和f_w分别为当前麻雀的适应度值、全局最佳适应度值和全局最差适应度值；ψ设为10^{-8}，避免分母为零。

1.3.3 SSA-XGBoost模型

利用SSA寻优能力强且收敛速度快的优势对XGBoost的核心参数：迭代次数、树深度与学习率等3个参数进行优化，即为SSA-XGBoost模型。

图2展示了SSA-XGBoost算法的具体流程。利用SSA全局寻优能力，在设置的搜索空间内对XGBoost的超参数（树深度、学习率）不断迭代优化，以构建最优SSA-XGBoost预测模型。本研究中，模型迭代次数范围为[10,50]，树的深度的寻优范围为[5,15]、学习率寻优范围为[0.01,0.3]。

1.4 模型预测结果评价

本研究采用准确率（Accuracy，A_{C}）评价模型的好坏[37]。准确率指所有样本中预测结果正确的比例。检验模型预测的结果除了用准确率之外，通常计算预测结果的精确率（Precision，P_{R}）

和召回率（Recall，R_E）[37]。精确率指在预测为同一类别的结果中预测正确的比例，召回率指相同类别被预测出来的比例。其计算公式分别如下：

$$A_C=(T_P+T_N)/(T_P+F_N+F_P+T_N) \quad (9)$$

$$P_R=T_P/(T_P+F_P) \quad (10)$$

$$R_E=T_P/(T_P+F_N) \quad (11)$$

式中：T_P为真正类的数量，即分类为正类，实际也是正类的样本数量；F_P为假正类的数量，即分类为正类，但实际是负类的样本数量；F_N为假负类的数量，即分类为负类，但实际是正类的样本数量；T_N为真负类的数量，即分类是负类，实际也是负类的样本数量。

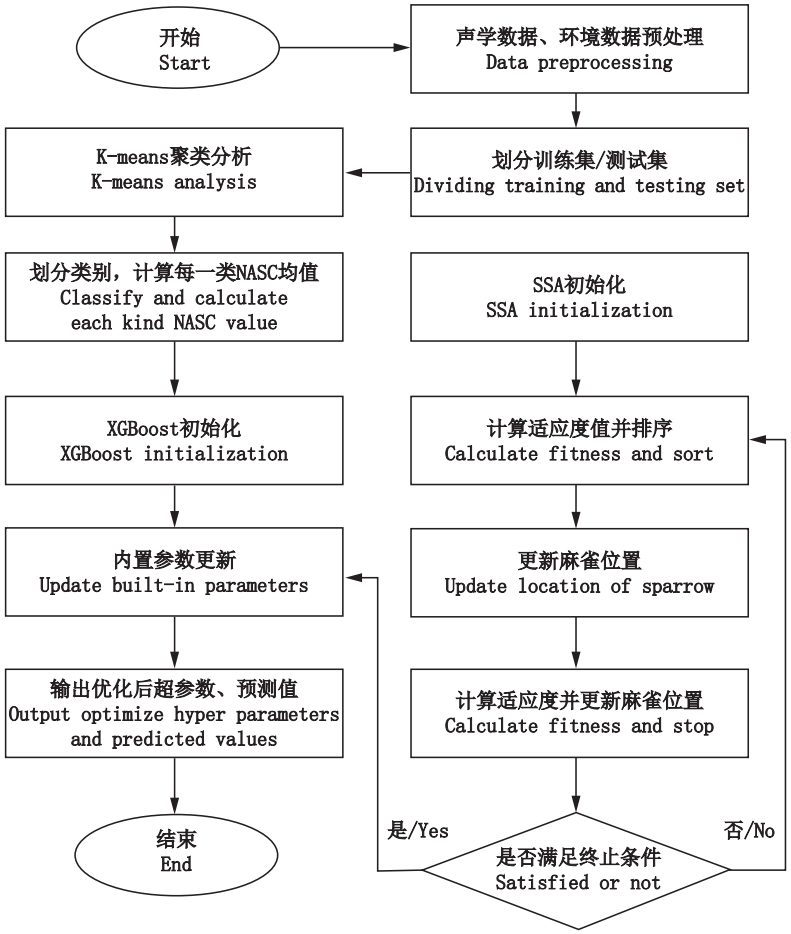

图2　SSA-XGBoost模型构建流程图

Fig. 2　Flow chart of SSA-XGBoost model

2　结果

2.1　SSA-XGBoost模型预测结果

研究共得到西南印度洋散射层的声学密度样本数据10 056个，基于K-means聚类分析，对所有样本进行分类，并统计了每一类别相应的各个环境因子均值与NASC的均值。结果表明（表2），在划分的10类样本中，占比最高的为第6类（C6）为21.87%，其NASC均值为21.88×10^4 m²/nmi²，占比最低的为第3类（C3）为0.86%，其NASC均值为9.11×10^4 m²/nmi²。NASC均值最高的为第7类（C7），为32.43×10^4 m²/nmi²，该类别占比为10.67%，NASC均值最低的为第9类（C9），为8.20×10^4 m²/nmi²，该类别占比为3.85%。

SSA-XGBoost模型的训练结果表明，该模型共迭代15次，麻雀的数量为6，最优超参数组合：树的深度为12，学习率为0.13。SSA-XGBoost模型的预测结果的混淆矩阵如图3所示，模型预测的准确率为80.51%，精确率76%，召回率78%。其中，该模型对第10类散射层预测的精确率最高为99.6%，对第3类的预测精确率最低为22.9%；对第6类预测的召回率最高为93.8%，第10类的召回率最低为63.1%。

表 2 K-means聚类分析结果
Tab. 2 Results of K-means clustering

类别 Class lable	类别占比 Proportion/ %	经度 Lon/ (°E)	纬度 Lat/ (°S)	叶绿素 *Chl.*a/ (mg/m^3)	溶解氧 DO/ (mmol/m^3)	pH	混合层深度 MLD/m	海表温度 SST/℃	涡动能 EKE/ (m^2/s^2)	光合有效辐射 PAR/ (W/m^2)	水深 Depth/ m	NASC均值 MeanNASC/ [×10^4(m^2/nmi^2)]
C1	2.96	69.96	44.26	0.34	283.45	8.04	90.71	10.79	0.20	22.61	3 893.39	17.22
C2	16.82	63.59	33.39	0.17	240.66	8.08	62.92	17.92	0.01	25.84	4 499.71	28.95
C3	0.86	73.51	48.55	0.83	337.46	8.04	80.07	3.28	0.03	17.63	1 812.83	9.11
C4	6.92	80.05	45.57	0.39	298.13	8.04	92.46	8.18	0.04	22.04	3 647.83	15.65
C5	9.95	68.54	40.37	0.38	256.09	8.06	220.50	13.47	0.03	18.30	4 282.11	31.44
C6	21.87	57.88	25.13	0.07	219.37	8.09	51.37	22.91	0.02	35.33	4 556.00	21.88
C7	10.67	67.77	39.06	0.34	256.34	8.05	48.63	15.09	0.02	38.79	4 313.76	32.43
C8	19.53	62.26	27.50	0.09	217.60	8.04	21.14	24.00	0.02	49.24	4 350.97	21.64
C9	3.85	72.36	49.11	0.35	331.72	8.03	70.15	3.91	0.01	22.35	1 442.84	8.20
C10	6.57	92.38	26.01	0.14	230.43	8.08	35.55	20.83	0.02	45.34	4 320.09	15.82

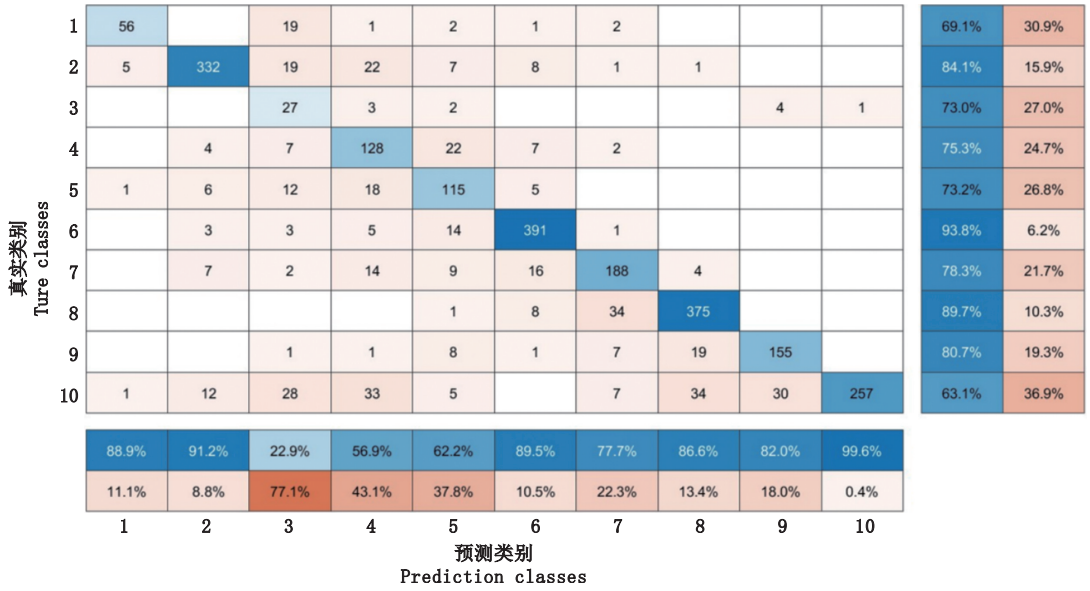

图 3 SSA-XGBoost模型训练结果混淆矩阵
Fig. 3 Confusion matrix of SSA-XGBoost model result

2.2 散射层资源密度预测及空间分布

基于训练好的SSA-XGBoost模型，本研究分别预测了西南印度洋2011年不同季节散射层的资源密度值。如图4所示，预测结果显示，散射层资源密度较高的海域集中在30°S~50°S之间海域，其中在夏季(1月)，该海域的西北区域的岛屿周边也存在密度较高的散射层。散射层密度较低的海域主要分布在50°S以南海域，以及该海域的东北区域。特别是在东北区域，随着季节的更替，该区域散射层的密度始终维持在较低水平。此外，在春季(10月)时散射层的高密度区分布更加集中，随着季节的推移，其高密度区域变得相对分散，其中秋季(4月)散射层的高密度区向南扩散较为明显。

为进一步分析西南印度洋海域散射层高密度区域的变动规律，提取了图4中NASC值最高(NASC>28×10^4 m^2/nmi^2)的2类散射层的空间分布重心，并计算得到了不同季节散射层重心的第一级标准差椭圆。如图5所示，随着季节的变动，散射层高密度区的重心由东南向西北方向移动，其中春季时期重心的纬度最大，冬季(7月)时重心的纬度最小。秋季时椭圆的旋转角度为90°(正东方向)，表明该季节散射层的重心整体呈东西方向分布，其他季节的旋转角度均大于90°，说明其他季节散射层的重心呈东南-西北方向分布。不同季节椭圆的长短轴的比值均大于2，说明重心点在东南-西北方向上的离散性大于东北-西南方向。

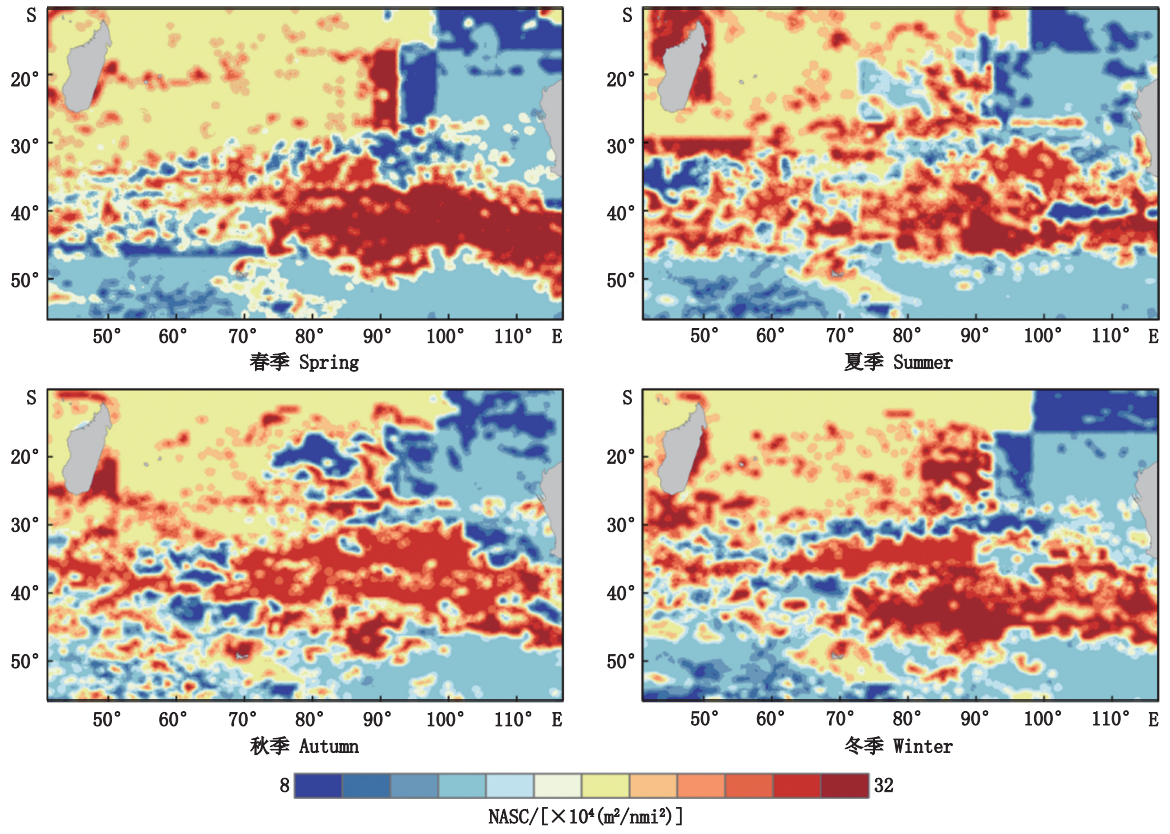

图4　2011年不同季节NASC预测结果

Fig. 4　Prediction results of different seasons in 2011

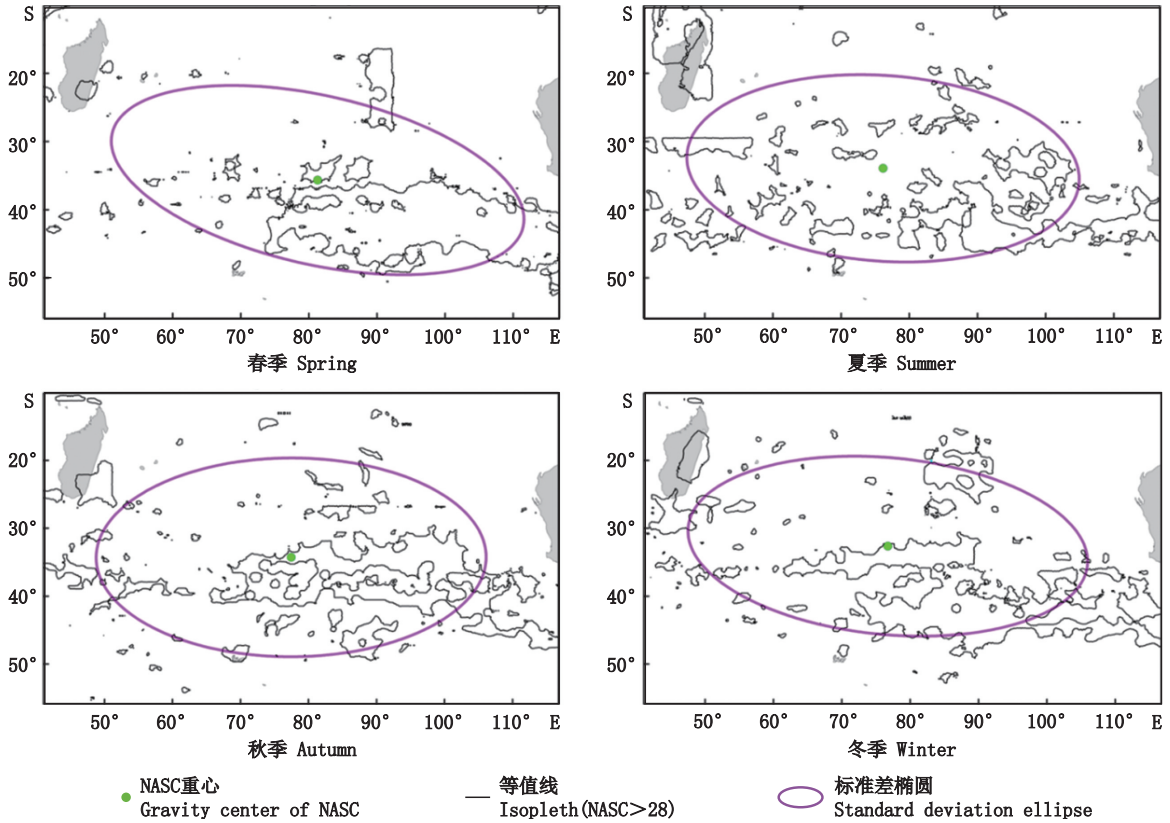

图5　2011年不同季节NASC预测结果高密度区域分布

Fig. 5　Spatial distribution of high density of predicted NASC of different seasons in 2011

3 讨论

3.1 NASC值在深海散射层研究中的应用

深海散射层中不同生物混栖，生物组成复杂，传统的资源评估方法难以实现散射层中生物量的计算。而声学调查通过测量散射层对声波的后向散射强度，在不完全明确生物种类的情况下，可以得到散射层相对资源丰度的声学指标，如S_v值、NASC值。其中本研究采用的NASC值在散射层的研究中得到了一定的应用。如BÉHAGLE等[17]研究发现西南印度洋散射层的NASC峰值出现在中纬度海域，且冬季相较于夏季NASC峰值的纬度会向北移动。ARIZA等[38]采用NASC值表征全球范围内散射层的资源丰度，预测了气候变暖的情境下，全球海洋的生物量将会出现下降。DIOGOUL等[39]发现塞内加尔大陆架海域散射层的NASC峰值与叶绿素峰值出现的规律相符，认为该海域散射层主要由浮游动物组成。

NASC值在无法厘清散射层中具体的生物种类时，可以有效表征散射层的相对资源密度，利于开展相关研究。值得注意的是，在采用NASC值时需要准确解读NASC值的意义。NASC值代表着单位积分水体内所有物体声散射强度。随着NASC值的增大，表示该单位积分水体中所有物体的总声散射强度增加。对散射层而言，生物类别复杂，全部的样本采集工作难以开展，且不同生物的单体目标强度不同，所以解释散射层NASC值存在一定难度。但仍可以从两个方面理解散射层NASC值的变化，首先，当积分水体内各物种的单体目标强度相近时，NASC值越大，表明该区域的生物量也越大；其次，当积分水体内各物种的单体目标强度差异较大时，NASC值越大，表示该区域的物种组成更为复杂，同时也可能存在某些物种的生物量较大[29]。尽管如此，需要明确的是仅凭NASC值的大小无法准确推断出散射层的生物量水平，只能代表相对水平。因此，散射层的研究需要综合声学调查与网具采样，并准确计算不同生物种类的目标强度。这对散射层的精准研究意义重大。

3.2 SSA−XGBoost模型的应用效果

深海散射层在大洋中广泛存在，与广阔的分布区域相比，沿调查船航迹连续记录的声学数据覆盖的采样区域仍然很小。因此，通过建立合适的模型预测整个海域散射层资源丰度，对深入理解散射层的资源分布尤为重要。XGBoost是机器学习中一种新兴的集成模型，它将多颗回归树组合起来形成一个性能更加强大的学习器，不仅对数据的拟合能力强于传统的线性回归，而且在模型构建过程中，将目标函数引入正则项，有效避免了变量较多而样本较少的资源丰度预测模型的过拟合[37]，但XGBoost也存在参数过多，难以找到较好的超参数组合的问题。麻雀搜索算法具有良好的全局搜索能力和局部寻优能力[36]，能够有效弥补XGBoost模型的这一缺陷。因此，SSA-XGBoost模型相比传统的资源丰度预测模型更适合高维海洋环境数据。

从本研究测试集的结果来看，SSA-Boost模型对K-means聚类结果中第3类的预测精确度最低（仅为22.9%），而将第3类较多的错误预测为第10类；同时对第10类的预测结果表明，仅将1个第10类的样本错误预测为第3类。但从整体来看，SSA-XGBoost模型的平均准确率、精确率、召回率分别为80.51%，76%和78%（表2），说明模型的精度较高、整体应用效果较好，只在部分类别的预测中效果较差。模型表现较差的原因有多种，其中不同类别之间区分度较小是导致模型训练效果不好、预测精度低的主要原因。此外，本研究共选取了8个环境因子和2个空间变量作为特征值对模型进行训练，尽管通过数据的标准化消除了各因子量级不同对训练模型产生的影响，但对于高维时空数据难以有效提取特征的问题并未筛选关键环境因子。将来可通过筛选关键环境因子减少输入数据的维度，降低模型训练难度和客观评价指标的冗余性与耦合性，提高模型精度。

此外，SSA-XGBoost模型具备良好的可解释性和稳健性。在可解释性方面，XGBoost内置了特征重要性评估工具，能够揭示模型对于不同特征的依赖程度。通过分析这些特征的重要性，能够直观地理解模型在决策过程中的关键因素。同时，XGBoost基于决策树的集成学习，其决策树结构相对简单，可以通过可视化工具清晰展示模型在每个决策节点上的决策逻辑，增强模型的可信度。在稳健性方面，SSA-XGBoost的预处理步骤有助于检测和处理时间序列中的异常值。首

先，通过降维和分解时间序列，SSA能够提高模型对异常值的鲁棒性，从而增强了整体模型的稳定性。其次，XGBoost通过正则化项控制模型的复杂度，有效防止过拟合，提高了模型对未知数据的泛化能力。在后续的应用中，可以考虑采用交叉验证，扩大数据量，通过在不同数据子集上训练和验证模型，进一步评估其性能和稳健性，确保模型在不同数据分布下都能表现良好。

3.3　散射层资源丰度的空间分布特征

本研究中散射层K-means聚类的结果表明，NASC均值最高的3类（NASC>28×10^4 m^2/nmi^2），其平均纬度分别为33.39°S，40.37°S，39.06°S。模型的预测结果表明，散射层的NASC高密度区基本位于30°S~50°S，说明实测数据与预测数据在空间分布上相吻合。BÉHAGLE等[17]和BOERSCH-SUPAN等[40]的研究也表明，西南印度洋海域散射层资源丰度最高的海区集中在亚热带海域，反映了该模型较好的应用效果。通过进一步分析预测数据的重心变动，发现散射层高密度区的重心由东南向西北方向移动，其中春季重心的纬度最大，冬季重心的纬度最小。重心点在东南-西北方向上的离散性大于东北-西南方向。散射层重心随季节的变动，可能是海洋环境的季节性差异引起的。研究[41]表明，初级生产力是影响海洋中层生物量的关键影响因素。在海洋上层，光照和温度随季节更替会出现较大的变化，由此驱动的海水理化环境的协同变化，进一步导致初级生产力会呈现出显著的季节特征，这一变化在温带和极地海区尤其明显[40]。另外在一些热带和亚热带大洋及边缘海，季风引发的上升流强度和营养盐水平的变化也会导致显著的初级生产力季节变化[40]。中尺度海洋动力过程，如中尺度涡，也被认为是驱动散射层空间分布变动的重要因素[42-43]。中尺度涡在变化的过程中，往往伴随着水体的移动，因此会直接将某一海区的浮游动植物直接输送到另一海区[44]。此外涡旋的类型决定了水体内营养盐和氧气等要素是否充足，从而改变浮游植物的生物量和分布，最终影响浮游动物的分布和群落结构[45-46]。所以，初级生产力作为散射层中许多生物的食物来源，这些变化都通过上行效应而进一步控制了散射层中物种的丰度、生物量和群落结构等表现出明显的季节和空间差异。

尽管通过相关模型的建立可以对散射层的资源丰度进行预测研究，但由于海洋环境变化过程的复杂性，预测结果往往会出现较大差异，如PROUD等[31]和ARIZA等[38]分别对全球尺度散射层资源丰度所作的预测分析，结果却呈现出两种完全相反的变化趋势。这种结果的差异表明了在目前环境变化的多重压力下，厘清引起散射层空间分布与资源变动机理的复杂性。同时，由于难以对散射层中的生物种类进行精准分类鉴定与目标强度建模，长期以来的研究多采用声散射强度表征其相对生物量，而无法准确评估绝对资源量。因此，将来散射层的研究需要在持续的声学观测和采样调查的基础上，发展更加精确的资源丰度预测模型，建立适用于散射层的声散射模型以评估其资源量，以期更加深入地认识散射层的时空分布特征及其与海洋环境的关系，探讨散射层在大洋碳、氮等生化要素汇集和释放过程中所产生的影响。

参考文献：

[1] DUVALL G E, CHRISTENSEN R J. Stratification of sound scatterers in the ocean [J]. The Journal of the Acoustical Society of America, 1946, 18(s1): 254.

[2] GIORLIG, DRAZEN J C, NEUHEIMER A B, et al. Deep sea animal density and size estimated using a Dual-frequency IDentificationSONar (DIDSON) offshore the island of Hawaii [J]. Progress in Oceanography, 2018, 160: 155-166.

[3] PRIOU P, NIKOLOPOULOS A, FLORES H, et al. Dense mesopelagic sound scattering layer and vertical segregation of pelagic organisms at the Arctic-Atlantic gateway during the midnight sun [J]. Progress in Oceanography, 2021, 196: 102611.

[4] POLIS G A, ANDERSON W B, HOLT R D. Toward an integration of landscape and food web ecology: the dynamics of spatially subsidized food webs [J]. Annual Review of Ecology and Systematics, 1997, 28: 289-316.

[5] KLEVJER T, MELLE W, KNUTSEN T, et al. Micronekton biomass distribution, improved estimates across four North Atlantic basins [J]. Deep Sea Research Part Ⅱ: Topical Studies in Oceanography, 2020, 180: 104691.

[6] DAWIDOWICZ P, PIJANOWSKA J, CIECHOMSKI K. Vertical migration of Chaoborus larvae is induced by the presence of fish [J]. Limnology and Oceanography, 1990, 35(7): 1631-1637.

[7] ROE H S J. Observations on the diurnal vertical migrations

of an oceanic animal community [J]. Marine Biology, 1974, 28(2): 99-113.

[8] KLOSER R J, RYAN T E, YOUNG J W, et al. Acoustic observations of micronekton fish on the scale of an ocean basin: potential and challenges [J]. ICES Journal of Marine Science, 2009, 66(6): 998-1006.

[9] 李玉昕, 杨颐华, 李志宽, 等. 南海深水散射层的实验研究[J]. 海洋学报, 1986, 8(1): 107-110.
LI Y X, YANG Y H, LI Z K, et al. Experimental study on the deep scattering layer in the South China Sea[J]. Acta Oceanologica Sinica, 1986, 8(1): 107-110.

[10] 裘辛方, 张仁和, 殷业. 黄海中部高频反向体积声散射的观察[J]. 声学技术, 1999, 18(2): 50-53.
QIU X F, ZHANG R H, YIN Y. Observations of high-frequency acoustic volume backscattering in the mid-Yellow Sea[J]. Technical Acoustics, 1999, 18(2): 50-53.

[11] 刘世刚, 汤勇, 陈国宝, 等. 南海深海声学散射层垂直分布和昼夜变化初步研究[J]. 海洋科学进展, 2015, 33(2): 173-181.
LIU S G, TANG Y, CHEN G B, et al. Vertical distribution and diurnal movement of the deep scattering layer inthe South China Sea [J]. Advances in Marine Science, 2015, 33(2): 173-181.

[12] 陈钊, 吕连港, 杨光兵, 等. 基于船载ADCP和LADCP观测的南海声散射层[J]. 海洋科学进展, 2016, 34(2): 240-249.
CHEN Z, LYU L G, YANG G B, et al. Research on sound scattering layer in the South China sea observed withship-board ADCP and LADCP [J]. Advances in Marine Science, 2016, 34(2): 240-249.

[13] 张超, 吕连港, 姜莹, 等. 基于多波束测深仪的西太平洋声散射层测量[J]. 海洋科学, 2018, 42(9): 1-9.
ZHAO C, LYU L G, JIANG Y, et al. Research on the sound-scattering layer in the Western Pacific observed witha multibeam sounding system [J]. Marine Sciences, 2018, 42(9): 1-9.

[14] 高爽, 杨光兵, 熊学军. 南海北部声散射季节变化[J]. 海岸工程, 2022, 41(2): 144-152.
GAO S, YANG G B, XIONG X J. Seasonal variation of sound scattering in the northern South China Sea [J]. Coastal Engineering, 2022, 41(2): 144-152.

[15] BERTRAND A, BARD F X, JOSSE E. Tuna food habits related to the micronekton distribution in French Polynesia [J]. Marine Biology, 2002, 140(5): 1023-1037.

[16] MOLINE M A, BENOIT-BIRD K, O'GORMAN D, et al. Integration of scientific echo sounders with an adaptable autonomous vehicle to extend our understanding of animals from the surface to the bathypelagic [J]. Journal of Atmospheric and Oceanic Technology, 2015, 32(11): 2173-2186.

[17] BÉHAGLE N, COTTÉ C, RYAN T E, et al. Acoustic micronektonic distribution is structured by macroscale oceanographic processes across 20-50° S latitudes in the South-Western Indian Ocean[J]. Deep Sea Research Part I: Oceanographic Research Papers, 2016, 110: 20-32.

[18] CASCÃO I, DOMOKOS R, LAMMERS M O, et al. Seamount effects on the diel vertical migration and spatial structure of micronekton [J]. Progress in Oceanography, 2019, 175: 1-13.

[19] BENOIT-BIRD K J, LAWSON G L. Ecological insights from pelagic habitats acquired using active acoustic techniques[J]. Annual Review of Marine Science, 2016, 8: 463-490.

[20] 孙昭, 李云, 江毓武, 等. 基于Stacking机器学习模型的南海北部海温预报[J]. 海洋预报, 2023, 40(1): 39-45.
SUN Z, LI Y, JIANG Y W, et al. Sea temperature forecast in the northern South China Sea base on Stacking machine learning model[J]. Marine Forecasts, 2023, 40(1): 39-45.

[21] 孟鑫垚, 刘焱雄, 陈义兰, 等. 基于空间特征的机器学习模型浅海水深反演[J]. 海洋科学进展, 2023, 41(3): 498-509.
MENG X Y, LIU Y X, CHEN Y L, et al. Spatial feature-based machine learning model for shallow water depth retrieval[J]. Advances in Marine Science, 2023, 41(3): 498-509.

[22] 李璠, 何丛颖, 李毅, 等. 基于机器学习的浮标实时自动化赤潮预警研究[J]. 中国环境监测, 2023, 39(4): 196-205.
LI F, HE C Y, LI Y, et al. Research on real-time automatic red tide warning of bathing buoy based on machine learning[J]. Environmental Monitoring in China, 2023, 39(4): 196-205.

[23] 谭雨欣, 田义超, 黄卓梅, 等. 北部湾茅尾海无瓣海桑红树林地上生物量反演——基于XGBoost机器学习算法[J]. 生态学报, 2023, 43(11): 4674-4688.
TAN Y X, TIAN Y C, HUANG Z M, et al. Aboveground biomass of *Sonneratia apetala* mangroves in Mawei Sea of Beibu Gulf based on XGBoost machine learning algorithm [J]. Acta Ecologica Sinica, 2023, 43(11): 4674-4688.

[24] 谢文鸿, 徐广珺, 董昌明. 基于ConvLSTM机器学习的风暴潮漫滩预报研究[J]. 大气科学学报, 2022, 45(5): 674-687.
XIE W H, XU G J, DONG C M. Research on storm surge floodplain prediction based on ConvLSTM machine learning [J]. Transactions of Atmospheric Sciences, 2022, 45(5): 674-687.

[25] 魏广恩. 北太平洋柔鱼渔场的时空分析与资源丰度的预测[D]. 上海: 上海海洋大学, 2018.
WEI G E. Spatial and temporal analysis of *Ommastrphes bartramii* fishing ground and its resource abundance prediction in the North Pacific Ocean [D]. Shanghai:

Shanghai Ocean University, 2018.

[26] 常亮，陈芳霖，陈新军，等. 基于BP神经网络的西北太平洋柔鱼资源丰度预测[J]. 上海海洋大学学报，2022，31(2)：524-533.
CHANG L, CHEN F L, CHEN X J, et al. Prediction of the CPUE of neon flying squid in the northwest Pacific Ocean based on back propagation neural network[J]. Journal of Shanghai Ocean University, 2022, 31(2): 524-533.

[27] 周茜涵，吴洽儿，周艳波，等. 基于优化灰色模型的南海鸢乌贼资源丰度预测[J]. 南方水产科学，2021，17(3)：1-7.
ZHOU X H, WU Q E, ZHOU Y B, et al. Prediction of abundance of *Sthenoteuthis oualaniensis* in South China Seabased on optimized grey system model[J]. South China Fisheries Science, 2021, 17(3): 1-7.

[28] HARIS K, KLOSER R J, RYAN T E, et al. Sounding out life in the deep using acoustic data from ships of opportunity [J]. Scientific Data, 2021, 8(1): 23.

[29] 屈泰春. 南极半岛北部和普里兹湾水域南极大磷虾的声学调查研究[D]. 大连：大连海洋大学，2014：20-32.
QU T C. Acoustic investigation of Antarctic krill in the northern Antarctic Peninsula and Pritz Bay[D]. Dalian: Dalian Ocean University, 2014: 20-32.

[30] MACLENNAN D N, FERNANDES P G, DALEN J. A consistent approach to definitions and symbols in fisheries acoustics[J]. ICES Journal of Marine Science, 2002, 59(2): 365-369.

[31] PROUD R, COX M J, BRIERLEY A S. Biogeography of the global ocean′s mesopelagic zone[J]. Current Biology, 2017, 27(1): 113-119.

[32] 吴凯，王晓琳，许怡，等. 中国大陆降水时空格局演变新事实[J]. 南水北调与水利科技，2017，15(3)：30-36.
WU K, WANG X L, XU Y, et al. New facts about evolution of spatial and temporal pattern of precipitation over Chinese mainland [J]. South-to-North Water Transfers & Water Science and Technology, 2017, 15(3): 30-36.

[33] CHEN T Q, GUESTRIN C. XGBoost: a scalable tree boosting system [C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco:ACM, 2016.

[34] XUE J K, SHEN B. A novel swarm intelligence optimization approach: sparrow search algorithm [J]. Systems Science & Control Engineering, 2020, 8(1): 22-34.

[35] MENG K, CHEN C, XIN B. MSSSA: a multi-strategy enhanced sparrow search algorithm for global optimization [J]. Frontiers of Information Technology & Electronic Engineering, 2022, 23(12): 1828-1847.

[36] 袁红春，高子玥，张天蛟. 基于改进的XGBoost模型预测南太平洋长鳍金枪鱼资源丰度[J]. 海洋湖沼通报，2022，44(2)：112-120.
YUAN H C, GAO Z Y, ZHANG T J. Prediction of albacore tuna abundance in south Pacific based on improved XGBoostmodel[J]. Transactions of Oceanology and Limnology, 2022, 44(2): 112-120.

[37] 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/OL]. 水产学报：1-13[2023-09-26]. http://kns. cnki. net/kcms/detail/31. 1283. S. 20221011. 1357. 012. html.
XIE M Y, LIN B, CHEN X J. Prediction on fishing ground of *Ommastrephes bartramii* in Northwest Pacific based on deep learning[J/OL]. Journal of Fisheries of China, 1-13 [2023-09-26]. http://kns. cnki. net/kcms/detail/31. 1283. S. 20221011. 1357. 012. html.

[38] ARIZA A, LENGAIGNE M, MENKES C, et al. Global decline of pelagic fauna in a warmer ocean [J]. Nature Climate Change, 2022, 12(10): 928-934.

[39] DIOGOUL N, BREHMER P, PERROT Y, et al. Fine-scale vertical structure of sound-scattering layers over an east border upwelling system and its relationship to pelagic habitat characteristics[J]. Ocean Science, 2020, 16(1): 65-81.

[40] BOERSCH-SUPAN P H, ROGERS A D, BRIERLEY A S. The distribution of pelagic sound scattering layers across the southwest Indian Ocean[J]. Deep Sea Research Part II: Topical Studies in Oceanography, 2017, 136: 108-121.

[41] 孙栋，王春生. 深远海浮游动物生态学研究进展[J]. 生态学报，2017，37(10)：3219-3231.
SUN D, WANG C S. A review of open ocean zooplankton ecology[J]. Acta Ecologica Sinica, 2017, 37(10): 3219-3231.

[42] JENNINGS S, MÉLIN F, BLANCHARD J L, et al. Global-scale predictions of community and ecosystem properties from simple ecological theory[J]. Proceedings of the Royal Society B: Biological Sciences, 2008, 275(1641): 1375-1383.

[43] GODØ O R, SAMUELSEN A, MACAULAY G J, et al. Mesoscale eddies are oases for higher trophic marine life [J]. PLoS One, 2012, 7(1): e30161.

[44] FENNELL S, ROSE G. Oceanographic influences on Deep Scattering Layers across the North Atlantic[J]. Deep Sea Research Part I: Oceanographic Research Papers, 2015, 105: 132-141.

[45] CONDIE S, CONDIE R. Retention of plankton within ocean eddies [J]. Global Ecology and Biogeography, 2016, 25(10): 1264-1277.

[46] HAUSS H, CHRISTIANSEN S, SCHÜTTEF, et al. Dead zone or oasis in the open ocean? Zooplankton distribution and migration in low-oxygen modewater eddies [J]. Biogeosciences, 2016, 13(6): 1977-1989.

Prediction on acoustic resource density of deep scattering layer of the southwestern Indian Ocean based on machine learning

WAN Shujie[1], CHEN Xinjun[1,2,3,4]

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

Abstract: Predicting the abundance and distribution of deep scattering layer is important to indicate the distribution of marine protected animals, important fishing grounds, and develop fishery resources into the scattering layer. This study used the Nautical Area Scattering Coefficient (NASC) as the resource density indicator of the scattering layer, and used K-means clustering and SSA-XGBoost model to predict the resource density of the scattering layer based on multiple environmental factors in the southwestern Indian Ocean. The results showed that the accuracy of the model prediction is 80.51%, the precision is 76%, and the recall is 78%. The sample data matches the high-density spatial distribution of the predicted data, and the application effect of the model is good. By predicting the density of the scattering layer in different seasons in 2011, it was found that the center of gravity in the high-density area of the scattering layer moved from southeast to northwest, with the latitude of the center of gravity being the largest in spring and the smallest in winter. The dispersion of the center of gravity in the southeast-northwest direction is greater than that in the northeast-southwest direction. This study can provide a new method for elucidating the distribution and resource variation patterns of scattering layers in larger spaces.

Key words: deep scattering layer; machine learning; acoustic resource density; southwestern Indian Ocean