

陈红梅. 基于 Sentinel-2 MSI 的养殖海湾富营养化反演研究[J]. 渔业研究, 2024, 46(6): 653 – 663.

基于 Sentinel-2 MSI 的养殖海湾富营养化反演研究

陈红梅

(福建省水产研究所, 福建 厦门 361013)

摘要: 【背景】水体富营养化会导致水生生态系统的失衡, 影响栖息生物的生存, 造成水产养殖经济损失。快速掌握重要养殖水域富营养化状态与变化趋势对海洋渔业发展与生态环境保护具有重大意义。【目的】本研究旨在建立适用于重要养殖海湾的富营养化指数反演模型, 为海洋环境保护、优化养殖规划布局等提供科学参考。【方法】基于实测水质参数和 Sentinel-2 MSI 卫星遥感影像数据, 筛选与富营养化指数的对数 $\lg(E)$ 相关性最高的 3 个波段组合作为特征波段输入至 CatBoost、BP 神经网络、随机森林 3 个机器学习模型, 对比 3 个模型的反演精度, 确定最优模型, 反演福建省诏安湾与东山湾 2022 年富营养化指数并进行时空特征分析。【结果】特征波段组合 $b3+b7$ 、 $b3-b12$ 、 $b3 \times b9$ 与具有较高的相关性, 作为反演富营养化指数的最佳波段组合; CatBoost 模型相比 BP 神经网络和随机森林有更高的反演精度, 决定系数 (R^2) 达到 0.90, 均方根误差 (RMSE) 为 5.67, 平均绝对百分比 (MAPE) 为 43.61%; 东山湾富营养化指数整体呈由湾顶朝湾外逐渐降低的趋势。诏安湾富营养化指数整体呈沿岸高、湾中低的趋势, 且有明显的冬季低、其他季节高的特点。【结论】东山湾、诏安湾富营养化时空分布有着明显差异。影响富营养化的因素有陆域污染物输入、养殖活动、地理特征、水文环境、天气变化等。本研究结果能快速、全面反映海湾水质时空变化趋势, 有助于定位污染源和敏感区, 可为环境评价、污染治理及养殖管理提供参考。

关键词: Sentinel-2; 富营养化; 水产养殖; CatBoost; BP 神经网络; 随机森林

中图分类号: S949; P237 **文献标识码:** A **文章编号:** 2096 – 9848 (2024) 06 – 0653 – 11

富营养化是一种氮、磷等营养物质含量过多所引起的水质污染现象。水体富营养化会导致水生生态系统的失衡, 影响栖息生物的生存, 造成水产养殖经济损失。快速掌握重要养殖水域富营养化状态与变化趋势对海洋渔业发展与生态环境保护具有重大意义。水质遥感反演监测具有覆盖范围广、快速、周期短和成本低的优势, 可动态、迅速地提供大面积海洋水质信息, 在一定程度上改善了以采样点数据代替附近水域水质参数浓度的不合理性, 可动态描述各指标的时间变化、空间分布趋势, 成为水质监测新的手段。已有众多学者利用 Landsat-8、TM、Sentinel-2、HJ-1、GOCI 等不同的卫星遥感

影像对总氮、总磷、化学需氧量、叶绿素 a 等水质参数的反演模型进行研究^[1-6]。传统的水质遥感方法分为经验法、半经验法与分析法^[7-8], 其中经验法和半经验法是通过分析实测水质参数与遥感波段之间的函数关系来构建一种线性或非线性反演模型的方法^[9], 在目前较为常用。但近岸二类水体具有更复杂的光学特性和影响因素^[10], 仅用简单的线性或非线性公式很难反映复杂的水体环境与遥感影像间的关系。近年来, 随着人工智能技术的发展, 越来越多的研究把机器学习理论融入水质遥感监测中。机器学习算法能够利用样本数据对现有的经验模型进行自主优化改进, 更好地挖掘数据之间潜在

收稿日期: 2024-11-03

修回日期: 2024-12-09

基金项目: 福建省属公益类科研院所基本科研专项 (2023R1012005); 福建省海洋服务与渔业高质量发展专项资金项目 (FJHY-YYKJ-2024-1-14、FJHY-YYKJ-2024-1-18-2)

第一作者: 陈红梅, 女, 助理研究员, 研究方向为海洋遥感与地理信息系统。E-mail: 563898690@qq.com

的联系,在处理非线性回归问题方面有明显优势^[11]。神经网络、随机森林、支持向量机、CatBoost 等机器学习方法已经被广泛用于水质参数的反演^[12]。当前许多研究基本集中于反演单个或多个水质参数,研究区域主要集中于内陆湖泊,较少对近岸二类水体进行富营养化状态的研究。

东山湾与诏安湾位于福建省南部,两湾经八尺门海峡水道相连^[13]。其中,东山湾呈不规则的梨形伸入陆地,湾顶有漳江入海,湾口朝南,两侧东山铜陵镇和漳浦古雷半岛遥相对望,构成经由口门东、西两水道及其上游5条指状槽沟连通的半封闭海湾。诏安湾位于东山湾西南面,海湾略呈南北伸展,湾顶有港口渡入海,沿岸另有公子店溪、林圻溪和石颯溪自北向南注入诏安湾,湾内海底宽浅平坦,湾口朝南,口门有城洲岛和西屿等岛屿屏障,口小腹大,受风浪影响小。东山湾与诏安湾是福建省重要的海水养殖基地。其中,东山湾以海水池塘养殖为主,主要分布于湾顶、湾西南部;其次为筏式养殖,分布于漳江口、东山岛北部至大霜岛以南海域;网箱养殖主要分布在东山湾口附近海域。诏安湾以筏式养殖为主,分布于诏安湾航道两侧;其次为沿岸池塘养殖;网箱养殖主要分布在田厝渔港

东南海域及西屿岛至东山岐下村海域。虽然工业发展、人口的增长以及水产养殖的快速发展促进了地方社会经济的发展,但在一定程度上产生了海洋环境质量下降的问题^[14]。同时,海洋环境质量下降也制约着海湾养殖的绿色健康发展。根据福建省生态环境厅发布的《福建省近岸海域水质状况排名(2022年1月—12月)》,在福建省重点港湾河口中,东山湾水质最优,而诏安湾位于倒数第三。诏安湾水环境状况备受当地政府部门和社会的高度关注^[15]。因此,本研究以诏安湾、东山湾为研究区域,使用 Sentinel-2 MSI 数据作为富营养化指数反演的光谱数据,结合实测水质参数和机器学习方法,旨在建立适用于福建省重要养殖海湾的富营养化指数反演模型,为海洋环境保护、优化养殖布局等提供科学参考。

1 材料与方法

1.1 实测水质参数及富营养化指数

于2022年9月13日至9月15日对诏安湾和东山湾进行实地考察,在69个站位进行水质样本采集(图1)。采样结束后,将水体样本带回实验室测定水质参数。检测项目包括化学需氧量、活性磷酸盐、亚硝酸盐-氮、硝酸盐-氮、氨-氮等。

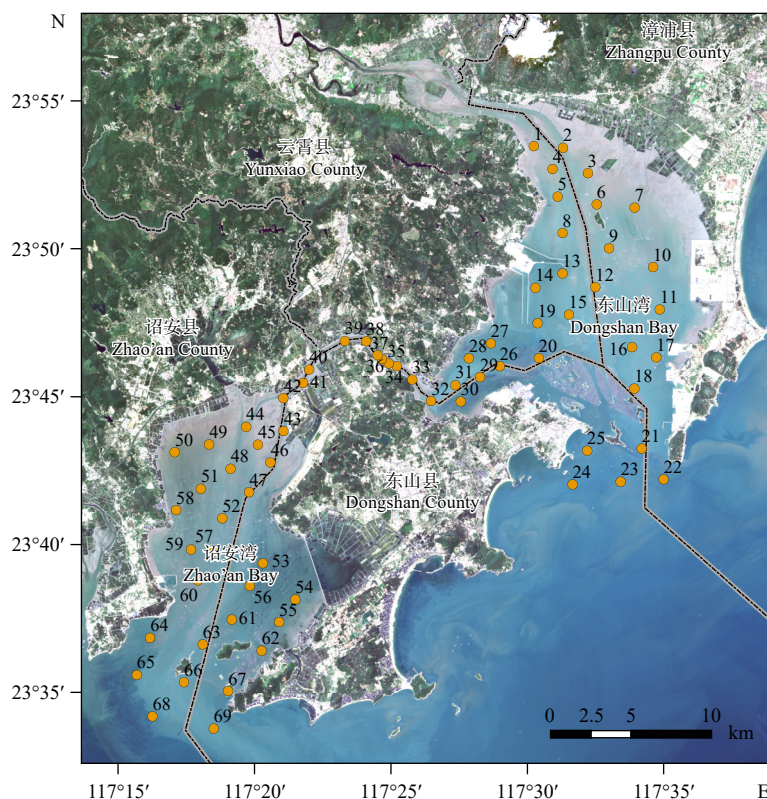


图1 研究区域及采样点分布

Fig. 1 Study area and distribution of sampling stations

本研究所选用的富营养化评价方法为富营养化指数法,由日本学者冈市友利^[16]于 1972 年提出,后经邹景忠等^[17]引进并对公式进行改进。自 2010 年起,国家海洋局发布的《中国海洋环境质量公报》正式利用该指数为近岸海水富营养化状态作出评价。

$$E = \frac{COD \times DIP \times DIN \times 10^6}{4\,500} \quad (1)$$

式(1)中: E 为富营养化指数; COD 为化学需氧量浓度 (mg/L); DIP 为活性磷酸盐浓度 (mg/L); DIN 为无机氮浓度 (mg/L)。其中 DIN 值为亚硝酸盐-氮、硝酸盐-氮、氨-氮值之和。2022 年 9 月 13 日至 9 月 15 日实测水质计算的 E 值结果如图 2 所示。

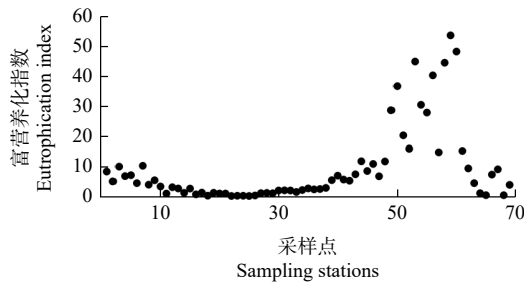


图 2 各采样点实测富营养化指数分布

Fig. 2 Distribution of measured eutrophication index at each sampling stations

1.2 Sentinel-2 MSI 数据

本研究使用的 Sentinel-2 MSI 数据来自欧空局网站 (<https://browser.dataspace.copernicus.eu/>)。Sentinel-2 由 2 颗卫星组成, Sentinel-2 A/B^[18] 星座分别于 2015 年与 2017 年发射, 卫星携带的 MSI 仪器能提供 13 个光谱波段, 主要用于监测土地环境、土地利用状况, 以及内陆水体与沿海水域, 还可用于监测山体滑坡、泥石流等自然灾害。Sentinel-2 MSI 的优势在于 10 m 的多光谱分辨率和较短的重返周期, 每颗卫星的运行周期为 10 d, 而当 2 组卫星同时运行时, 运行周期则为 5 d, 大幅提升了 Sentinel-2 MSI 的对地观测能力^[19], 因此其被广泛用于水质的监测。研究共选取 6 景 2022 年研究区域的无云影像进行研究, 6 景影像分布于 4 个季节, 其中 9 月 15 日的影像用于富营养化指数反演建模, 其余影像用于研究区域内富营养化指数的时空变化分析。所选影像具体信息如表 1 所示。

1.3 反演模型

1.3.1 CatBoost

CatBoost 是一种新型的梯度提升 (GBDT) 算

法^[20]。该算法通过排序梯度提升降低了训练样本的噪声数据, 克服了传统 GBDT 算法中过拟合与梯度偏差的问题^[21-22], 相比 XGBoost 和 LightBGM, CatBoost 提高了算法精度和模型的泛化能力^[23]。公式^[24]为:

$$x_k^i = \frac{\sum_{x_{i,j} \in E_k} \{x_j^i = x_k^i\} \times y_i + ap}{\sum_{x_{i,j} \in E_k} \{x_j^i = x_k^i\} + a} \quad (2)$$

式(2)中: x_k^i 为第 k 个样本的第 i 个样本特征; x_j^i 为第 k 个样本之前第 j 个样本的第 i 个类别特征; y_j 为第 j 个样本的标签值; E_k 为随机序列中在第 k 个样本之前的数据集; p 为添加的先验值和大于 0 的权重系数; a 为通常大于 0 的权重系数。

表 1 研究选用的 Sentinel-2 MSI 影像

Tab. 1 Sentinel-2 MSI images selected for the study

序号 Serial number	季节 Seasons	时间 Time	数据级别 Data level
1	冬	2022-01-03	L2A
2	春	2022-04-08	L2A
3	夏	2022-07-22	L2A
4	夏	2022-09-15	L2A
5	秋	2022-11-19	L2A
6	冬	2022-12-24	L2A

在训练时, CatBoost 使用多行串连的方式集合多个学习器, 每经过一轮学习即更新一次权重, 从而逐步降低噪声点带来的偏差^[25]。使用 CatBoost 模型训练需要为几个基础参数进行赋值: iterations 为模型训练时需要的迭代次数; learning_rate 为学习率, 用于控制每次迭代时更新权重的幅度; l2_leaf_reg 为 l2 正则化系数, 可以防止模型过拟合, 使模型拥有良好的泛化能力; max_depth 为最大树深, 用于控制决策树的复杂程度, 同样可以避免模型的过拟合; loss_function 为损失函数, 用于衡量预测值与实际值之间的差异。

1.3.2 BP 神经网络

BP 神经网络是一种多层前馈的神经网络, 是目前最常用的机器学习算法^[26]。该模型将学习过程分为向前传输和误差反向传播两个部分^[27]。在训练过程中, 输入信号经过隐藏层节点, 经过非线性变换产生输出值。若输出值与期望值不符, 则开始误差的反向传播过程, 将误差均分给各个节点, 从而调整各节点权重, 反复循环, 最终确定误差最小的阈值。BP 神经网络需要指定的参数如下:

hidden_layer_sizes 用于指定隐藏层的大小与每个节点中神经元的数量；max_iter 为最大迭代次数；solver 为优化算法，用于求解权重参数的优化问题；activation 为激活函数，用于在神经网络的隐藏层和输出层中引入非线性变换；learning_rate_init 为学习率。

1.3.3 随机森林

随机森林是一种集成学习算法^[28]，它是以决策树作为弱分类器的集成分类器^[29]，利用 Bootstrap 重抽样方法在原始样本中抽取多个样本，对每个样本进行建模并预测，最终投票得出预测结果^[30]。

随机森林可以在保持高维特征输入的基础上，评估每个模型参数的重要性^[28]，这种特性使其在遥感信息提取方面得到广泛应用^[12]。随机森林模型需考虑的参数较少，主要取决于决策树个数和最大特征数，在本研究中需要指定的参数除了常规的学习率、最大树深外，还有 min_samples_leaf 和 max_features，他们分别代表叶节点上的最小样本数和最大特征数。

1.4 反演方法

本文技术路线包括数据预处理、训练集构建、模型训练、时空分析（图 3）。

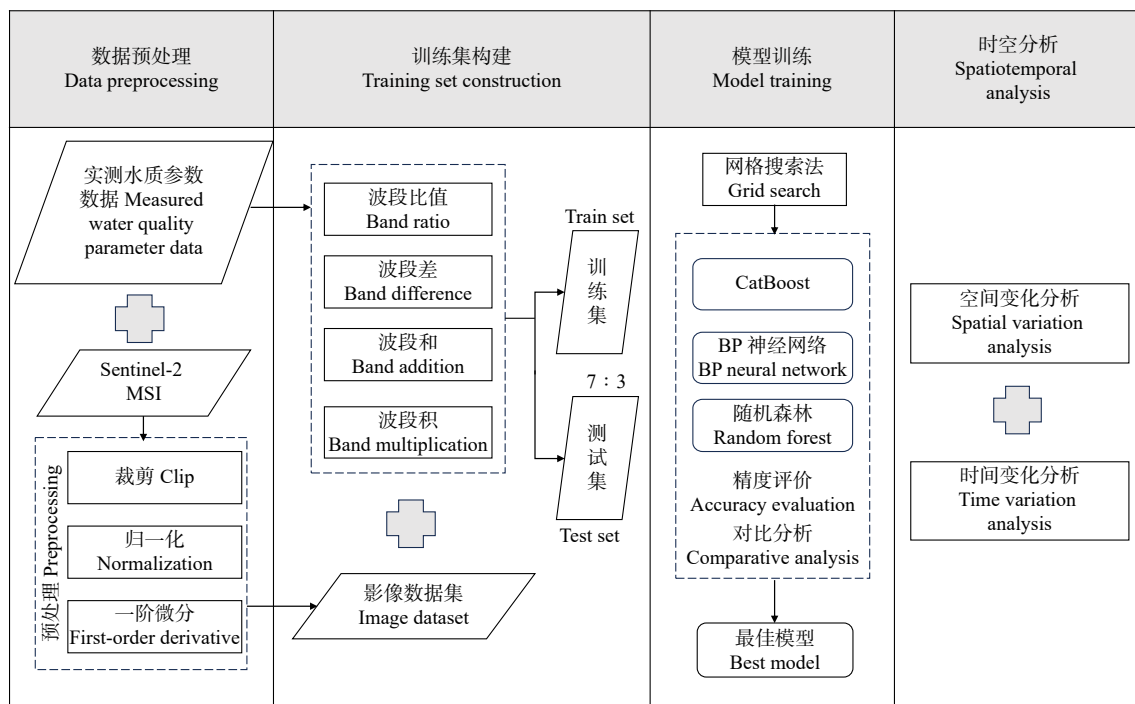


图 3 研究技术路线图

Fig. 3 Research technology roadmap

1.4.1 数据预处理

欧空局网站下载的 L2A 级数据已进行大气校正与几何校正处理，因此在本研究中，只需进行图像裁剪与后续的预处理工作。同时，由于研究区域水面上存在部分网箱养殖和滩涂，因此使用 McFeeters S K^[31] 提出的归一化水体指数（Normalized difference water index, NDWI）对研究区域水体进行提取，以剔除水上网箱养殖、滩涂对建模及后续分析的影响。NDWI 公式为：

$$NDWI = \frac{\rho_{Green} - \rho_{NIR}}{\rho_{Green} + \rho_{NIR}} \quad (3)$$

式（3）中： ρ_{Green} 、 ρ_{NIR} 分别为绿光波段和近

红外波段，分别代表 Sentinel-2 MSI 影像中的第三波段和第八波段。

由于光谱信息易受到外界干扰，水体间相似的光谱特性可能导致信息冗余。为了最大程度地获取有效光谱信息^[32]，进一步提升不同水体之间光谱信息的差异性，提高富营养化指数反演的精度，本研究选择对 Sentinel-2 MSI 图像进行一阶微分处理。一阶微分可以去除部分线性或接近线性的背景、噪声光谱对目标光谱的影响^[33]。光谱一阶微分公式为：

$$R(\lambda') = \frac{R(\lambda_{i+1}) - R(\lambda_{i-1})}{\lambda_{i+1} - \lambda_{i-1}} \quad (4)$$

式(4)中: $R(\lambda)$ 为第 i 波段对应的一阶微分值; $R(\lambda_{i+1})$ 、 $R(\lambda_{i-1})$ 分别为第 $i+1$ 波段和第 $i-1$ 波段的原始值; λ_{i+1} 、 λ_{i-1} 分别为第 $i+1$ 波段和第 $i-1$ 波段的中心波长。

1.4.2 训练集构建

将数据集按 7:3、8:2、6:4 划分为训练集和测试集是常见的做法, 其中 7:3 的划分适合大多数中等规模数据集^[34]。为了能更好地提升模型的性能, 防止模型过拟合或欠拟合, 本文在建模过程中将实测的 69 个点以 7:3 划分为训练集和测试集, 并保证划分后的训练集和测试集在每个富营养化指数区间段内都占有一定比例。同时, 对实测得到的富营养化指数取以 10 为底的对数形式 $\lg(E)$, 消除异方差对参数估计的影响^[35], 使处理后的数据更为平稳。通常情况下, 随着遥感影像的波段增多, 光谱信息会更加丰富, 但数据冗余度也会增加^[36], 因此需要构建最佳波段组合形式, 以减轻数据的冗余, 加强光谱特征。本研究基于一阶微分后的 Sentinel-2 MSI 共 12 个波段构建 4 种波段组合形式: 波段比值、波段和、波段差和波段积。通过遍历各个波段组合与 $\lg(E)$, 计算 Pearson 相关系数, 选择相关系数最高的 3 个组合作为输入模型的 3 个特征波段。

1.4.3 模型训练

分别对 CatBoost、BP 神经网络、随机森林 3 种模型进行训练, 经精度评价对比分析后, 选择最优模型。在模型的参数确定上, 通过网格搜索法^[37]确定 3 个模型的最优参数(表 2)。在模型的精度评价方面, 本研究选取决定系数 (R^2), 均方根误差 (Root mean squared error, RMSE) 和平均绝对百分比误差 (Mean absolute percentage error, MAPE)

作为评判模型精度的指标。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}} \quad (6)$$

$$MAPE = \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times \frac{100}{n} \quad (7)$$

式(5)~(7)中: y_i 表示实测富营养化指数值; $\hat{y}_i$ 表示模型预测富营养化指数值; $\bar{y}$ 表示实测富营养化指数值的平均数。

2 结果与分析

2.1 最优波段组合

部分波段组合与 $\lg(E)$ 的相关系数如表 3 所示。结果表明, 波段组合 b3+b7、b3-b12、b3×b9 与 $\lg(E)$ 具有较高的相关性。因此, 本文选用 b3+b7、b3-b12、b3×b9 共 3 个波段组合作为模型的输入因子, 以 $\lg(E)$ 作为模型的输出因子, 最后取反对数得到研究区域的富营养化指数 E 。

表 2 所选模型重要参数最优值

Tab. 2 The optimal value of important parameters of the model

模型 Models	最优参数 Optimal parameters
Catboost	iterations=1000, learning_rate=0.01, l2_leaf_reg=1, max_depth=3, loss_function='RMSE'
BP 神经网络 BP neural network	hidden_layer_sizes=(5,), max_iter=1000, solver='lbfgs', activation='relu', learning_rate_init=0.01
随机森林 Random forest	n_estimators=1000, random_state=42, max_depth=4, min_samples_leaf=1, max_features=3

表 3 部分波段组合与 $\lg(E)$ 的相关系数

Tab. 3 Correlation coefficient of some band combinations with $\lg(E)$

组合 Combinations	相关系数 Correlation coefficient	组合 Combinations	相关系数 Correlation coefficient
波段比值 Band ratio	b9/b6 -0.70	波段和 Band summation	b3+b7 0.82
	b12/b6 0.69		b3+b8 0.78
	b3/b4 -0.68		b2+b3 0.78
波段差 Band difference	b3-b12 0.79	波段积 Band multiplication	b3×b9 -0.81
	b2-b12 0.75		b3×b2 -0.77
	b3-b5 0.74		b2×b9 -0.69

2.2 模型反演精度

CatBoost、BP 神经网络、随机森林 3 种模型的反演结果如图 4 所示。CatBoost 模型的 R^2 达到

0.90, RMSE 为 5.67, MAPE 为 43.61%; BP 神经网络模型的 R^2 只有 0.75, RMSE 为 7.08, MAPE 为 42.81%; 随机森林模型的 R^2 达到 0.87, RMSE 为

7.04, $MAPE$ 为 49.41%。

从图4可以看出, CatBoost 模型的反演精度最高, 反演结果更趋近于1条直线, 对比其余2个模型, CatBoost 模型的预测值在每个富营养化指数区间内的分布更加紧凑。BP神经网络反演结果中 $MAPE$ 值最低, 但其在富营养化指数为10左右的区间内比较分散, 导致 R^2 较低。随机森林模型在

富营养化指数为5左右的区间内预测值偏高, 虽然 R^2 能达到 0.87, 但随着富营养化指数值的增加, 预测误差也会逐渐扩大, 这使得该模型达到了三者中最高的 $MAPE$ 值。从预测结果被分类至正确的区间来看, CatBoost 模型的正确率达到 90%, BP神经网络的正确率为 66%, 随机森林的正确率为 86%。

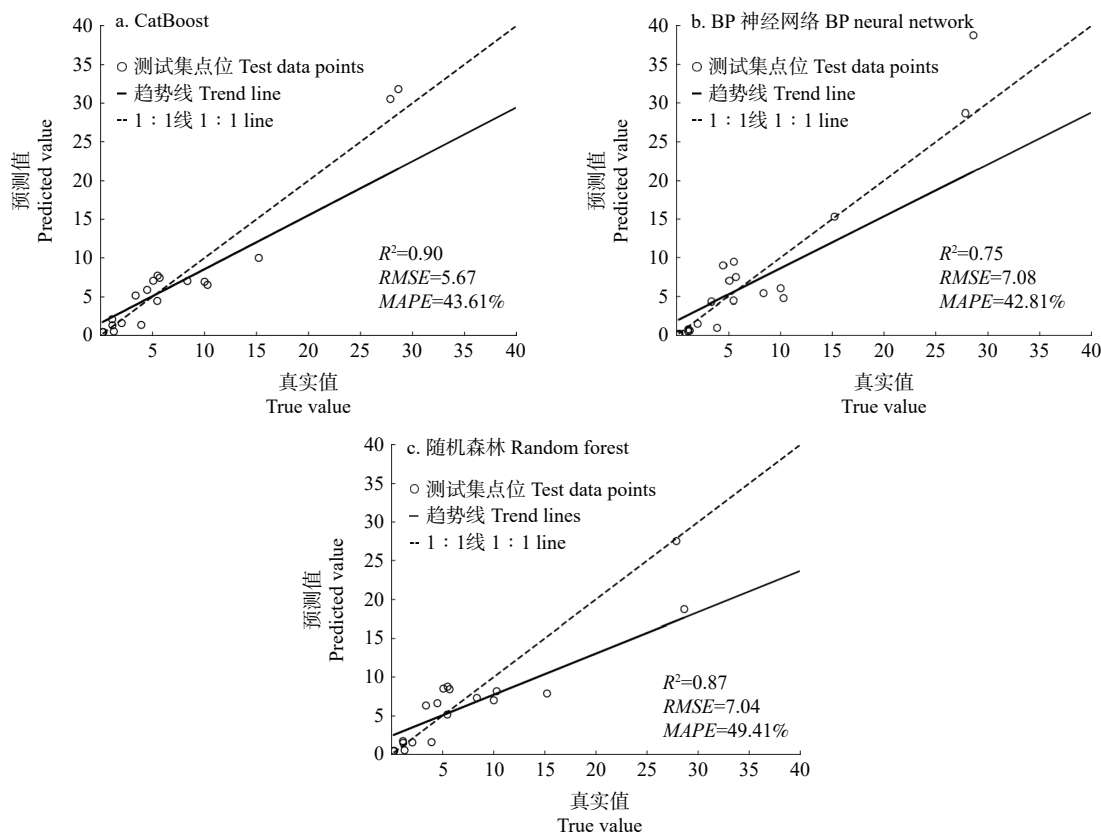


图4 不同模型富营养化指数反演精度评价

Fig. 4 Evaluation of inversion accuracy of eutrophication index by different models

本研究用于输入模型的特征波段组合由 b3、b7、b9、b12 组成, 这 4 个波段代表 Sentinel-2 MSI 的绿、近红外、短波红外波段等, 这与水质变化对光谱的敏感波段一致。相比 BP 神经网络和随机森林, CatBoost 能在设定同样数量模型参数的情况下更好地模拟沿海水域水体的富营养化指数。综合考虑不同模型的预测精度、分类结果及普适性, 本研究最终选用 CatBoost 模型作为福建省诏安湾与东山湾富营养化指数反演模型。

2.3 富营养化时空变化分析

本文采用构建的 CatBoost 模型分别对东山湾、诏安湾 2022 年 1 月 3 日、4 月 8 日、7 月 22 日、9 月

15 日、11 月 19 日、12 月 24 日六景 Sentinel-2 MSI 影像进行富营养化指数反演, 结果见图 5。

从富营养化指数反演结果图可知, 东山湾不同时期绝大部分区域富营养化指数较低, 水质状况较好。富营养化指数高值区主要分布于东山湾湾顶及其西南部八尺门口附近海域、漳州核电取排水导堤附近海域, 整体呈现由湾顶朝湾外逐渐降低的趋势 (图 5)。分析原因可能为漳江流入东山湾带来了大量陆源污染物; 而八尺门附近、漳州核电取排水导堤附近海域因受取排水导堤建设影响, 水动力条件变差, 导致水中富营养物质富集, 不易扩散, 造成水体环境质量相对湾口较差。

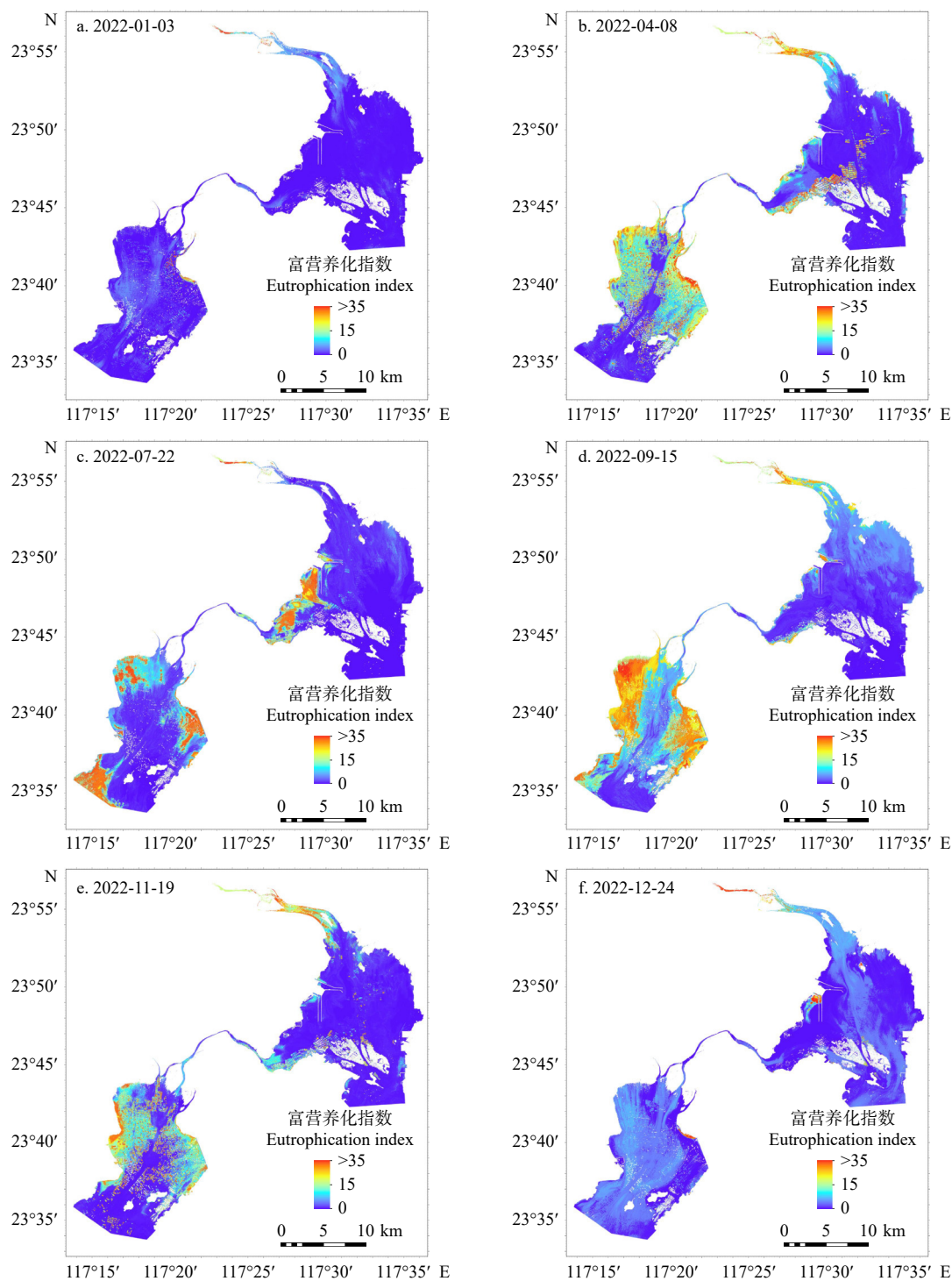


图 5 不同时期 Sentinel-2 影像富营养化指数反演结果

Fig. 5 Inversion results of eutrophication index in Sentinel-2 images in different periods

诏安湾 2022 年 1 月 3 日、12 月 24 日（冬季）富营养化指数整体均呈现低值，其余季节富营养化指数相对较高（图 5）。分析原因可能是冬季因降水较少，通过径流输入到海湾中的陆源污染物减少，两岸养殖池塘、养殖生产减少，营养盐源强减小；同时，冬季盛行的西风带有助于带动湾内海水

流动，扩散条件好；而进入春季以后，受降雨影响，陆源污染物通过径流输入到海湾中的量增加，加上养殖生产增多，海湾富营养化程度开始上升。从空间分布上看，诏安湾富营养化指数总体呈岸高、湾中低的趋势。主要原因为诏安湾沿岸受陆源工业、生活污染影响大，同时近岸池塘养殖、工厂

化养殖养殖尾水排放也增加了营养盐的输入,以及近岸水深较浅,加上滩涂、海上养殖设施使水体交换能力变弱,一定程度上造成了营养盐富集;而海湾中部由于水动力条件较好,污染物扩散能力较强,因此整体上诏安湾沿岸侧富营养化指数较海湾中部高。

3 结论

本文基于实测水质参数和 Sentinel-2 MSI 卫星遥感影像数据,筛选出与富营养化指数的对数 $\lg(E)$ 相关性最高的 3 个波段组合,作为特征波段输入至 CatBoost、BP 神经网络、随机森林 3 个机器学习模型。对比 3 个模型的反演精度,确定最优模型为 CatBoost。应用 CatBoost 模型反演福建省诏安湾和东山湾 2022 年 6 个不同时期富营养化指数,并进行时空特征分析。

1) 结果表明 $b3+b7$ 、 $b3-b12$ 、 $b3 \times b9$ 三波段组合与 $\lg(E)$ 具有最高的相关性,相关性均在 0.8 左右,以此作为 CatBoost、BP 神经网络、随机森林 3 个模型的特征波段进行输入。

2) 对比不同模型的反演结果,可以得出 CatBoost 能更好地模拟沿海水域水体的富营养化指数,其反演的福建省诏安湾和东山湾富营养化指数 R^2 可达到 0.90, $RMSE$ 为 5.67, $MAPE$ 为 43.61%,综合表现在三个模型中最优。

3) 东山湾富营养化指数整体呈现由湾顶朝湾外逐渐降低的趋势。诏安湾富营养化指数整体呈沿岸高、湾中低的趋势,且有明显的冬季低、其他季节高的特点。

东山湾、诏安湾富营养化时空分布有着明显差异。影响海湾富营养化的因素有陆域污染物输入、养殖活动、地理特征、水文环境、天气变化等。本研究结果能全面、快速反映海湾水质时空变化趋势,有助于定位污染源和敏感区,为环境评价、污染治理及养殖管理提供参考。

参考文献 (References):

[1] 金惠淑, 鱼京善, 孙文超, 等. 基于 GOCI 遥感数据的湖泊富营养化监测研究 [J]. 北京师范大学学报 (自然科学版), 2013, 49 (2-3): 271-274.
Jin H S, Yu J S, Sun W C, *et al.* Monitoring lake eutrophication by GOCI remote sensing [J]. Journal of Beijing Normal University (Natural Science), 2013,

49(2-3): 271-274.
[2] 刘静, 况润元, 李建新, 等. 基于实测数据的鄱阳湖总氮、总磷遥感反演模型研究 [J]. 西南农业学报, 2020, 33 (9): 2088-2094.
Liu J, Kuang R Y, Li J X, *et al.* Inversion model of TN, TP concentration based on measured spectral reflectance data in Poyang Lake [J]. Southwest China Journal of Agricultural Sciences, 2020, 33(9): 2088-2094.
[3] 乔平林, 张继贤, 林宗坚. 石羊河流域水质环境遥感监测评价研究 [J]. 国土资源遥感, 2003, 15 (4): 39-41, 45.
Qiao P L, Zhang J X, Lin Z J. The application of remote sensing technique to monitoring and evaluating water pollution in the Shiyang River valley [J]. Remote Sensing for Natural Resources, 2003, 15(4): 39-41, 45.
[4] 王凤霞, 夏卓异, 郭雨辉, 等. 基于 GEE 的中国南海水质反演与富营养化评价 [J]. 中国环境科学, 2022, 42 (2): 826-833.
Wang F X, Xia Z Y, Guo Y H, *et al.* Water quality inversion and eutrophication assessment of the South China Sea based on GEE [J]. China Environmental Science, 2022, 42(2): 826-833.
[5] Cao J J, Wen X P, Luo D Y, *et al.* Study on water quality inversion model of Dianchi Lake based on Landsat 8 data [J]. Journal of Spectroscopy, 2022, 2022(1): 3341713.
[6] Yao Y J, Zhu L, Wu C Q, *et al.* Water quality remote sensing monitoring research in China based on the HJ-1 satellite data [C]//Institute of Electrical and Electronics Engineers. Proceedings of 2010 IEEE International Geoscience and Remote Sensing Symposium. Honolulu: IEEE, 2010: 620-623.
[7] Yang H B, Du Y, Zhao H L, *et al.* Water quality Chl-*a* inversion based on spatio-temporal fusion and convolutional neural network [J]. Remote Sensing, 2022, 14(5): 1267.
[8] 李云梅, 黄家柱, 韦玉春, 等. 用分析模型方法反演水体叶绿素的浓度 [J]. 遥感学报, 2006, 10 (2): 169-175.
Li Y M, Huang J Z, Wei Y C, *et al.* Inversing chlorophyll concentration of Taihu Lake by analytic model [J]. Journal of Remote Sensing, 2006, 10(2): 169-175.
[9] Yang H B, Kong J L, Hu H H, *et al.* A review of remote sensing for water quality retrieval: progress and chal-

- lenges[J]. *Remote Sensing*, 2022, 14(8): 1770.
- [10] 李云梅, 赵焕, 毕顺, 等. 基于水体光学分类的二类水体水环境参数遥感监测进展 [J]. *遥感学报*, 2022, 26 (1): 19–31.
- Li Y M, Zhao H, Bi S, *et al.* Research progress of remote sensing monitoring of case II water environmental parameters based on water optical classification[J]. *National Remote Sensing Bulletin*, 2022, 26(1): 19–31.
- [11] Chen J, de Hoogh K, Gulliver J, *et al.* A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide[J]. *Environment International*, 2019, 130: 104934.
- [12] 卢雪梅, 苏华. 基于 OLCI 数据的福建近海悬浮物浓度遥感反演 [J]. *环境科学学报*, 2020, 40 (8): 2819–2827.
- Lu X M, Su H. Retrieving total suspended matter concentration in Fujian coastal waters using OLCI data[J]. *Acta Scientiae Circumstantiae*, 2020, 40(8): 2819–2827.
- [13] 中国海湾志编纂委员会. 中国海湾志 [M]. 北京: 海洋出版社, 1993.
- Editing Board of Records of Bays in China. Records of bays in China[M]. Beijing: China Ocean Press, 1993.
- [14] 郑惠东. 东山湾水产养殖承载力指标体系的构建与评价 [J]. *渔业研究*, 2019, 41 (5): 393–398.
- Zheng H D. Constructing and evaluating an index system for aquaculture carrying capacity in Dongshan Bay [J]. *Journal of Fisheries Research*, 2019, 41(5): 393–398.
- [15] 吴舜聪. 诏安湾水环境现状调查及整治对策 [J]. *化学工程与装备*, 2020 (5): 304–306.
- Wu S C. Investigation and remedial measures for the current situation of water environment in Zhao'an Bay[J]. *Chemical Engineering & Equipment*, 2020(5): 304–306.
- [16] 岡市友利. 浅海の汚染と赤潮の発生 [R]. 东京: 水产学研究叢書 (日本水产资源保护协会), 1972.
- Gangshiyouli. The pollution of Shallow Sea and the occurrence of red tide, mechanism of the occurrence of red tide in Inner Bay[R]. Tokyo: Japan Aquatic Resources Protection Association, 1972.
- [17] 邹景忠, 董丽萍, 秦保平. 渤海湾富营养化和赤潮问题的初步探讨 [J]. *海洋环境科学*, 1983, 2 (2): 41–54.
- Zou J Z, Dong L P, Qin B P. Preliminary study on eutrophication and red tide issues in Bohai Bay[J]. *Marine Environmental Science*, 1983, 2(2): 41–54.
- [18] Drusch M, Del Bello U, Carlier S, *et al.* Sentinel-2: ESA's optical high-resolution mission for GMES operational services[J]. *Remote Sensing of Environment*, 2012, 120: 25–36.
- [19] Caballero I, Steinmetz F, Navarro G. Evaluation of the first year of operational Sentinel-2A data for retrieval of suspended solids in medium- to high-turbidity waters[J]. *Remote Sensing*, 2018, 10(7): 982.
- [20] Prokhorenkova L, Gusev G, Vorobev A, *et al.* CatBoost: unbiased boosting with categorical features[C]//NeurIPS. Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal: Curran Associates Inc., 2018: 6639–6649.
- [21] 孔瑞瑶, 谢涛, 马明, 等. CatBoost 模型在水深反演中的应用 [J]. *测绘通报*, 2022 (7): 33–37.
- Kong R Y, Xie T, Ma M, *et al.* Application of CatBoost model in water depth inversion[J]. *Bulletin of Surveying and Mapping*, 2022(7): 33–37.
- [22] 李永康, 王新军, 马燕飞, 等. 基于 Catboost 的 AMSR-2 半经验地表温度降尺度 [J]. *干旱区研究*, 2021, 38 (6): 1637–1649.
- Li Y K, Wang X J, Ma Y F, *et al.* Downscaling land surface temperature through AMSR-2 passive microwave observations by Catboost semiempirical algorithms[J]. *Arid Zone Research*, 2021, 38(6): 1637–1649.
- [23] Bentéjac C, Csörgő A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms[J]. *Artificial Intelligence Review*, 2021, 54(3): 1937–1967.
- [24] Zhang Y X, Zhao Z G, Zheng J H. CatBoost: a new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China[J]. *Journal of Hydrology*, 2020, 588: 125087.
- [25] 陈点点, 陈芸芝, 冯险峰, 等. 基于超参数优化 CatBoost 算法的河流悬浮物浓度遥感反演 [J]. *地球信息科学学报*, 2022, 24 (4): 780–791.
- Chen D D, Chen Y Z, Feng X F, *et al.* Retrieving suspended matter concentration in rivers based on hyperparameter optimized CatBoost algorithm[J]. *Journal of Geo-Information Science*, 2022, 24(4): 780–791.
- [26] Verrelst J, Muñoz J, Alonso L, *et al.* Machine learning

- regression algorithms for biophysical parameter retrieval: opportunities for Sentinel-2 and -3[J]. *Remote Sensing of Environment*, 2012, 118: 127–139.
- [27] 郭文, 马梦梦, 孙培彦. 基于 Sentinel-1A 数据和 BP 神经网络的裸露地表土壤含水量反演研究 [J]. *中国农村水利水电*, 2023 (1): 89–94.
- Guo W, Ma M M, Sun P Y. Research on the soil water content inversion based on Sentinel-1A data and BP neural network[J]. *China Rural Water and Hydropower*, 2023(1): 89–94.
- [28] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5–32.
- [29] 杨庆振, 郭敏, 范新成. 基于随机森林算法的高光谱遥感作物分类 [J]. *测绘与空间地理信息*, 2023, 46 (4): 149–151, 154.
- Yang Q Z, Guo M, Fan X C. Hyper-spectral remote sensing crop classification based on random forest algorithm[J]. *Geomatics & Spatial Information Technology*, 2023, 46(4): 149–151, 154.
- [30] 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 technologies on random forests[J]. *Statistics & Information Forum*, 2011, 26(3): 32–38.
- [31] McFeeters S K. The use of the normalized difference water index (NDWI) in the delineation of open water features[J]. *International Journal of Remote Sensing*, 1996, 17(7): 1425–1432.
- [32] 梁金晨, 江晓东, 杨沈斌, 等. 基于高光谱遥感数据的水稻叶温反演 [J]. *南方农业学报*, 2020, 51 (1): 230–236.
- Liang J C, Jiang X D, Yang S B, *et al.* Rice leaf temperature inversion based on hyperspectral remote sensing data[J]. *Journal of Southern Agriculture*, 2020, 51(1): 230–236.
- [33] 庞博, 李玉霞, 童玲. 基于半分析模型的叶绿素 *a* 浓度估算 [J]. *成都理工大学学报 (自然科学版)*, 2011, 38 (1): 54–58.
- Pang B, Li Y X, Tong L. Estimation of chlorophyll-*a* concentration in Kuncheng Lake of Jiangsu based on a semi-analytical model[J]. *Journal of Chengdu University of Technology (Science & Technology Edition)*, 2011, 38(1): 54–58.
- [34] Xu L, Hou L, Zhu Z Y, *et al.* Mid-term prediction of electrical energy consumption for crude oil pipelines using a hybrid algorithm of support vector machine and genetic algorithm[J]. *Energy*, 2021, 222: 119955.
- [35] 曾伟生, 唐守正. 非线性模型对数回归的偏差校正及与加权回归的对比分析 [J]. *林业科学研究*, 2011, 24 (2): 137–143.
- Zeng W S, Tang S Z. Bias correction in logarithmic regression and comparison with weighted regression for non-linear models[J]. *Forest Research*, 2011, 24(2): 137–143.
- [36] 王芳. 最佳波段组合的典型地物信息提取 [J]. *航天返回与遥感*, 2022, 43 (2): 82–91.
- Wang F. Best band combination of Chinese high-resolution satellite image information extraction[J]. *Spacecraft Recovery & Remote Sensing*, 2022, 43(2): 82–91.
- [37] 陈宇, 白征东. 基于多极值网格搜索法的快速影像匹配技术研究 [J]. *测绘通报*, 2010 (4): 28–30.
- Chen Y, Bai Z D. Research on fast image matching based on multi-extreme grid search method[J]. *Bulletin of Surveying and Mapping*, 2010(4): 28–30.

Inversion on eutrophication in breeding bay based on Sentinel-2 MSI data

CHEN Hongmei

(Fisheries Research Institute of Fujian, Xiamen 361006, China)

Abstract: [Background] The eutrophication of water bodies will lead to an imbalance in aquatic ecosystems, affect the survival of habitats, and cause economic losses of aquaculture. Rapidly assessing the eutrophic conditions and evolving trends of key aquaculture waters are highly significant for the advancement of marine fisheries and the protection of the ecological environment. [Objective] This study aims to establish an inversion model of the eutrophic index applicable to important aquaculture bays to provide scientific reference for marine environmental protection and aquaculture planning. [Methods] The study utilized the measured water quality parameters and Sentinel-2 MSI satellite remote sensing image data, screened out the three band combinations with the highest correlation with the logarithmic $\lg(E)$ of the eutrophication index as characteristic bands. These bands were then inputted into three machine learning models: CatBoost, BP neural network, and random forest. The study compared the inversion accuracy of these models to determine the optimal one. Subsequently, the study used this model to invert the eutrophication status of the waters in the Zhao'an Bay and Dongshan Bay areas of Fujian Province in 2022, and analyzed the spatial and temporal characteristics. [Results] The results showed that $b3+b7$, $b3-b12$, and $b3 \times b9$ were the best band combinations for inversion of eutrophication index, with correlations all around 0.8. The CatBoost model had higher inversion accuracy than BP neural network and random forest. The coefficient of determination R^2 reached 0.90, the root mean square error (RMSE) was 5.67, and the mean absolute percentage (MAPE) was 43.61%, respectively, the eutrophication index of Dongshan Bay showed a gradually decreasing trend from the top of the bay towards the outside. The eutrophication index of Zhao'an Bay showed a trend of high along the coast and low in the middle of the bay, with obvious characteristics of low in winter and high in other seasons. [Conclusion] There are significant differences in the spatial and temporal distribution of eutrophication between Dongshan Bay and Zhao'an Bay. The factors that affect eutrophication include input of terrestrial pollutants, seasonal changes in aquaculture, geographical features, hydrological environment, weather changes, etc. These findings provide methodological support for large-scale, rapid, and convenient eutrophication monitoring, and provide reference for environmental assessment, governance, and aquaculture management.

Key words: Sentinel-2; eutrophication; aquaculture; CatBoost; BP neural network; random forest