

# 一种融合多波段比值法和随机森林算法的Sentinel-2绿潮信息提取方法

周红武<sup>1</sup>, 宁珊<sup>1</sup>, 李达<sup>1</sup>, 潘锡山<sup>2</sup>, 李乔<sup>1</sup>,  
赵敏<sup>1</sup>, 唐潇<sup>1</sup>

(1.南京师范大学海洋科学与工程学院, 江苏南京 210023; 2.江苏省海涂研究中心, 江苏南京 210023)

**摘要:** 遥感卫星可以高效、大面积地监测海洋绿潮的发生及发展情况, 而绿潮的精准识别提取尤为重要, 对于维护海洋生态环境安全具有重大意义。本研究提出一种融合多波段比值法和随机森林算法 (random forest, RF) 的绿潮信息提取方法, 基于高分辨率的 Sentinel-2 遥感影像, 以苏北近岸海域为研究区, 以绿潮和非绿潮两类地物为识别目标进行精准识别分类, 并与最大似然法 (maximum likelihood classification, MLC)、RF 和支持向量机 (support vector machine, SVM) 以及融合多波段比值法的 MLC 和 SVM 等五种模型进行对比。结果表明: 归一化植被指数 (normalized difference vegetation index, *NDVI*) 和漂浮藻类指数 (floating algae index, *FAI*) 能够有效提高绿潮信息提取的准确性, 融合这两种指数的 RF 模型分类结果 Kappa 系数为 0.98, 总体精度为 99.92%, 高于其他五种模型。研究结果可为海洋环境生态及生物地球化学循环的相关研究与治理提供理论与技术支撑。

**关键词:** 黄海绿潮; Sentinel-2; *NDVI*; *FAI*; 随机森林

中图分类号: X87 文献标识码: A 文章编号: 1007-6336(2024)01-0119-11

## A Sentinel-2 green tide information extraction method incorporating multi-band ratio method and Random Forest algorithm

ZHOU Hongwu<sup>1</sup>, NING Shan<sup>1</sup>, LI Da<sup>1</sup>, PAN Xishan<sup>2</sup>, LI Qiao<sup>1</sup>,  
ZHAO Min<sup>1</sup>, TANG Xiao<sup>1</sup>

(1.School of Marine Science and Engineering, Nanjing Normal University, Nanjing 210023, China; 2.Tidal Flat Research Center of Jiangsu Province, Nanjing 210023, China)

**Abstract:** Remote sensing satellites enable an efficient and extensive monitoring of the occurrence and development of marine green tides, while the accurate identification and extraction of green tides are significantly important for the protection of the marine ecological environment. The present study proposed a green tide information extraction method with the combinations of the multi-band ratio method and RF algorithm. Based on the high-resolution Sentinel-2 imagery, we distinguished between the green tide and non-green tide features in northern Jiangsu province offshore area. Five other models were compared with our developed identification avenue, including MLC, RF, SVM, MLC with multi-band ratio method, and SVM with multi-band ratio method. The results showed that the two vegetation indices, *NDVI* and *FAI* could markedly improve the accuracy of green tide information extraction. The RF model, which combined these two vegetation indices, provided a Kappa coefficient of 0.98 and an overall accuracy of 99.92%, outperforming the

收稿日期: 2023-05-05, 修订日期: 2023-08-01

基金项目: 江苏省海洋科技创新专项项目 (JSZRHYKJ202206); 江苏省生态环境监测科研基金项目 (2329)

作者简介: 周红武 (1995-), 男, 山东济宁人, 硕士研究生, 主要研究领域为海洋遥感, E-mail: zhouhwu@njnu.edu.cn

通信作者: 李达 (1983-), 男, 江西赣州人, 博士, 副教授, 主要从事海洋生态环境研究, E-mail: lida@nju.edu.cn

other five models. These findings of our study could offer, to some extent, the theoretical and operational supports for the future related research and the scientific management of the ecological and biogeochemical processes in marine environments.

**Key words:** green tide; Sentinel-2; *NDVI*; *FAI*; random forest

绿潮 (green tides) 是海水中大型绿藻爆发性增殖或高度聚集引起水体变色的一类生态异常现象, 大规模绿潮会竞争性消耗水体中的营养物质和氧气, 造成海洋生物死亡, 海草床生境退化等危害<sup>[1]</sup>。近年来, 生活污水和工业废物的排放导致世界范围内水体富营养化问题日益严重, 加之全球气候变化, 绿潮在全球大规模频繁爆发, 受到国内外海洋生态学家、藻类学家的广泛关注<sup>[2]</sup>。2007 年, 我国黄海海域首次出现大规模绿潮爆发的现象, 山东、江苏两省合计打捞绿藻高达 6000 吨<sup>[1]</sup>, 每年 5—8 月该海域绿潮都会反复爆发, 至今已持续 15 年, 对该海域的经济和社会发展造成了巨大影响<sup>[2]</sup>。作为世界上最大规模的绿潮, 目前黄海绿潮被认为是该海域最严重的海洋生态灾害之一<sup>[3]</sup>。黄海绿潮爆发之初, 对绿潮发生机制以及防控经验不足, 多以绿潮爆发后的拦截打捞和岸滩清理为主, 防控成本较高。但由于水文环境复杂且绿潮规模过大, 只能对重要海域、近岸海域以及风景旅游区进行防控, 布控海域仅占黄海绿潮入侵海域的一小部分<sup>[4]</sup>。因此, 绿潮的动态监测及精准防控, 对于黄海绿潮防灾减灾工作尤为重要。

卫星遥感技术凭借其大尺度、多空间分辨率、多光谱和高时效性的优点, 成为绿潮动态监测的重要手段。近年来, 国内外学者基于多源遥感影像和漂浮绿藻光谱特性提取影像中的绿潮信息, 并取得一系列的研究成果。常见的绿潮提取算法多为多波段比值法, 包括归一化植被指数 (normalized difference vegetation index, *NDVI*)、增强型植被指数 (enhanced vegetation index, *EVI*)、漂浮藻类指数 (floating algae index, *FAI*) 和针对 *GOCI* 卫星的漂浮绿藻指数 (index of floating green algae for *GOCI*, *IGAG*) 等。王国伟等<sup>[5]</sup>利用 *MODIS* 数据计算长时间序列的 *NDVI*, 以确定黄海绿潮的 *NDVI* 阈值; 邢前国等<sup>[6]</sup>基于 *Landsat-TM* 及 *MODIS* 数据, 采用动态 *NDVI* 阈值的方法在不同水体中提取绿潮斑块信息, 并提出高浑浊

水体中使用高分辨率影像提取绿潮信息更加有效; Son 等<sup>[7]</sup>使用 *GOCI* 数据对 2011 年黄海绿潮信息进行提取时, 发现 *EVI* 算法也能取得较好效果; 袁超等<sup>[8]</sup>结合 *NDVI* 阈值分割和目视解译法对 Sentinel-2 高分辨率影像提取绿潮信息, 并对苏北浅滩 2018 年早期绿潮漂移路径进行监测分析; 南佛罗里达大学的 Hu<sup>[9]</sup>认为 *NDVI* 对环境变化及观测条件较为敏感, 针对复杂的海洋大气环境提出了 *FAI*, 并基于 *MODIS* 数据在黄海海域验证了其可靠性。综合上述国内外研究成果, 针对遥感卫星影像的黄海绿潮信息提取问题, 大多学者使用的影像数据为中低分辨率数据 (如 *MODIS*、*Landsat-TM*、*GOCI*), 且研究区多集中于山东海域, 而在黄海绿潮首次出现的苏北海域, 由于绿潮初期斑块面积较小, 中低分辨率卫星难以对其进行准确提取<sup>[8]</sup>。在光谱指数的选取方面, *NDVI* 结合红外和近红外波段的信息, 可以有效提取遥感影像中的植被信息<sup>[5]</sup>, 且该指数在绿潮提取研究被广泛应用。同时, *FAI* 对环境观测条件 (如气溶胶类型和厚度、观测角度、太阳耀光等) 的敏感度较低, 并且可以穿透薄云, 使用该指数可以较容易地识别和圈定各种海洋水域的浮游藻类<sup>[9]</sup>。目前大多研究仅使用一种指数进行绿潮信息的提取分析, Narin 等<sup>[10]</sup>发现单一指数受环境变化影响较大, 多指数融合可有效提高地物分类精度。因此, 针对复杂多变的海洋环境, 单一指数高精度识别绿潮的能力需进一步验证和改进。

传统的多波段比值法虽然在绿潮信息提取研究中被广泛应用, 但是由于遥感影像空间尺度较大, 像元分割阈值在单景影像中也可能发生较大变化<sup>[11]</sup>, 仅利用多波段比值法难以满足绿潮动态监测的需求。目前, 机器学习算法逐渐应用到水生植被遥感监测领域。Oiry 等<sup>[12]</sup>基于 Sentinel-2 数据, 使用随机森林算法 (random forest, RF) 提取了法国 26 个海湾的潮间带植被,

其中包括底栖微藻和少量绿藻。李敬虎等<sup>[11]</sup>提出一种基于深度学习的无人机影像夜光藻赤潮提取方法。张佩莹等<sup>[13]</sup>则是将 RF 分类法与水体植被指数结合, 基于 Sentinel-2 数据提取了水库水生植被, 并与最大似然法 (maximum likelihood classification, MLC) 和支持向量机 (support vector machine, SVM) 进行分类精度对比; 发现相比于 MLC 和 SVM 方法, 与水体植被指数相融合的 RF 方法在水生植被分类中具有更好的适用性。而机器学习算法在海洋绿潮信息提取中应用较少, 蔡晓晴<sup>[14]</sup>将机器学习算法应用到黄海绿潮遥感监测中, 基于 SVM 对传统绿潮信息提取方法进行改进, 实现了 GOCI 数据中不同藻种与海水的精准识别。RF 作为一种集成学习算法, 具有对多元共线性不敏感, 无须对数据进行归一化处理, 结果对缺失和非平衡的数据有较好的容忍度等优点<sup>[15]</sup>。如将 RF 与水体植被指数相融合, 可以解决多波段比值法阈值不易确定的缺点<sup>[13]</sup>, 同时相较传统机器学习算法能够充分利用遥感图像的光谱信息<sup>[16]</sup>, 大幅提升绿潮的识别能力。

本研究采用空间分辨率为 10 m 的 Sentinel-2 数据, 研究针对高分辨率多光谱遥感数据的苏北近岸海域绿潮信息提取问题。我们结合多波

段比值法和 RF 算法, 构建一种融合  $NDVI/FAI$  的 RF 模型, 对黄海绿潮空间分布进行高精度提取。将提出的融合 RF 方法, 与 MLC、RF 和 SVM 以及融合多波段比值法的 MLC 和 SVM 等五种模型进行比较, 以评价该方法的优劣。本研究结果可为早期绿潮在苏北海域的动态监测提供一种技术手段。

## 1 数据与方法

### 1.1 研究区

近 15 年船舶和卫星监测发现黄海绿潮多在 4 月中下旬零星出现在苏北辐射沙洲, 5 月中下旬开始在苏北浅滩以及外海海域形成斑块或条带状绿潮, 受潮汐、流场和海表风场作用不断聚集增殖, 6 月底在山东半岛海域大面积爆发并登岸, 7—8 月绿潮面积逐渐减少直至消亡<sup>[17]</sup>。本研究选取南黄海北部的苏北近岸海域 (33.33°N—35.24°N, 119.75°E—122.02°E) 作为研究区 (图 1)。它的西北部毗邻山东日照海域, 东北部与苏北浅滩特有地形地貌——辐射沙洲相连。苏北近岸海域适宜的气候、水温以及盐度使其成为中国最大的条斑紫菜养殖基地, 而养殖梗绳上的附生绿藻在富营养化的养殖海区爆发生长被认为是黄海绿潮的来源之一<sup>[18]</sup>。

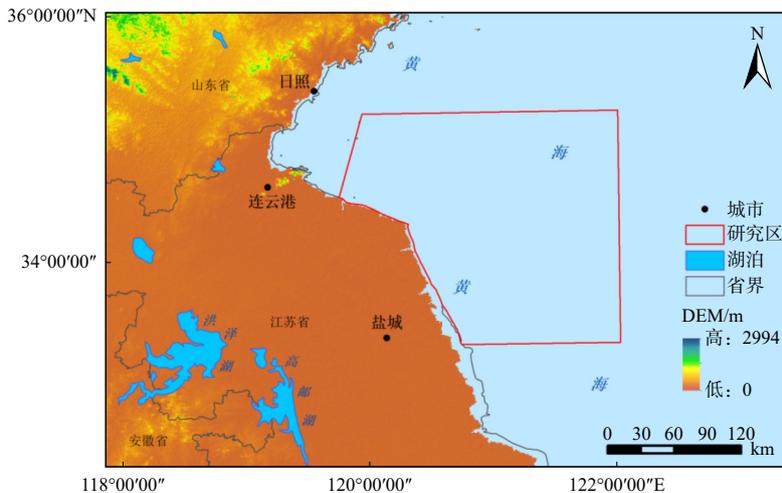


图 1 研究区位置

Fig. 1 Location of the study area

### 1.2 研究数据

本研究使用的 Sentinel-2 卫星 Level-1C 级数

据产品, 可通过欧空局哥白尼数据中心 (<https://scihub.copernicus.eu/dhus/>) 免费获取。Sentinel-2

卫星是欧洲航天局哥白尼计划发射的高分辨率多光谱成像卫星。Sentinel-2 卫星高度为 786 km, 幅宽达 290 km, 覆盖 13 个光谱波段, 空间分辨率为 10 m、20 m 和 60 m(表 1)。由于 Sentinel-2 数据在红边范围包含三个波段, 其对植被信息的识别监测方面优于其他光学影像数据。而 Sentinel-2 卫星的高时空分辨率特征为研究南黄海早期绿潮在苏北近岸海域的分布、发生和发展规律提供数据支撑。根据《中国海洋灾害公报》记载, 近 15 年黄海绿潮在卫星影像中首次出现多集中于 5 月中下旬, 结合 Sentinel-2 遥感影像云量覆盖情况, 本研究选取 2021 年 6 月 7 日的 4 景 Sentinel-2 遥感影像(影像编号: T51STT、T51STU、T51SUT、T51SUU), 影像镶嵌之后的空间范围基本覆盖早期绿潮发生的苏北近岸海域。

表 1 Sentinel-2 遥感影像基本参数

Tab.1 Basic parameters of Sentinel-2 remote sensing data

波段	光谱区间	中心波长 /nm	带宽 /nm	空间分辨率 /m
B1	可见光(VIS)	443	20	60
B2		490	65	10
B3		560	35	10
B4		665	30	10
B5	近红外(NIR)	705	15	20
B6		740	15	20
B7		783	20	20
B8		842	115	10
B8a		865	20	20
B9		945	20	60
B10		短波红外(SWIR)	1375	30
B11	1610		90	20
B12	2190		180	20

2021 年 6 月 5 日—10 日, 我们在南黄海海域出海同步执行绿潮跟踪任务, 船舶航行区域为 34.37°N—35.24°N, 120.31°E—120.90°E 的海域。我们在航区布放漂浮跟踪器并进行船舶跟踪。由于 Sentinel-2 卫星重访周期为 5 天, 考虑到绿潮的漂移速度, 本研究选取 6 月 6 日—6 月 8 日的航次同步调查数据进行现场验证。另外, 在样本数据的选择上, 我们结合现场调查数据, 在 4 景 Sentinel-2 遥感影像的基础上, 通过目视

解译来获得“真值”, 选择 150 处的非绿潮、120 处的绿潮作为训练样本, 并随机选择绿潮和非绿潮共计 300 处作为验证数据, 以此比较不同模型的提取精度。

### 1.3 研究方法

#### 1.3.1 归一化植被指数

本研究基于 Sentinel-2 数据, 利用多波段比值法和随机森林算法构建黄海绿潮信息提取的方法。通过主成分变换提取影像特征值至少几个主分量中, 然后利用多波段比值法计算 *NDVI* 和 *FAI* 两种指数, 在此基础上融合随机森林算法构建 RF 分类算法进行 Sentinel-2 影像中黄海绿潮的识别提取<sup>[19]</sup>。

归一化植被指数<sup>[5]</sup>通过测量近红外(植被强烈反射)和红光(植被吸收)之间的差异来量化植被, 具体表达式为:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

式中: *NIR* 为近红外波段的反射率; *RED* 为红光波段的反射率。

在本研究中, 影像均经过辐射定标、大气校正、重采样等预处理, 根据黄海绿潮光谱特征, *RED* 采用 Sentinel-2 的 B4 波段, *NIR* 采用 B8 波段。

#### 1.3.2 浮游藻类指数

浮游藻类指数<sup>[20]</sup>利用多光谱影像的红光、近红外、短波红外波段进行组合运算, 突出了浮游藻类与水体以及其他地物之间的差异, 具体公式为:

$$FAI = R_{NIR} - R'_{NIR} \quad (2)$$

$$R'_{NIR} = R_{RED} + (R_{SWIR} - R_{RED}) \times \frac{\lambda_{NIR} - \lambda_{RED}}{\lambda_{SWIR} - \lambda_{RED}} \quad (3)$$

式中: *R<sub>RED</sub>* 为红光波段的反射率; *R<sub>NIR</sub>* 为近红外波段的反射率; *R<sub>SWIR</sub>* 为短波红外波段的反射率;  $\lambda_{RED}$  为红光波段的中心波长;  $\lambda_{NIR}$  为近红外波段的中心波长;  $\lambda_{SWIR}$  为短波红外波段的中心波长; *R'<sub>NIR</sub>* 为插值反射率, 即红光和短波红外波段在近红外波段处采用线性内插方式得到的反射率信息。

在本研究中, 根据前人对黄海绿潮光谱特征的研究, *R<sub>RED</sub>* 采用 Sentinel-2 的 B4 波段, *R<sub>NIR</sub>* 采用 B8 波段, *R<sub>SWIR</sub>* 采用 B10 波段;  $\lambda_{RED}=665 \text{ nm}$ ,  $\lambda_{NIR}=842 \text{ nm}$ ,  $\lambda_{SWIR}=1375 \text{ nm}$ 。

### 1.3.3 主成分分析

主成分分析 (principal component analysis, PCA) 是一种统计方法, 为简化数据的处理和分析, 用主成分分析可将不同事物间相互融合和冗余的特性转化为少数几个不相关的新特性。在遥感影像处理过程中, 保证信息容量最大化的前提下压缩参与分类的数据量 (即减少光谱影像波段), 可减少分类所需时间, 提高分类速度<sup>[21]</sup>。由于本研究选取的 Sentinel-2 影像包含 13 个波段, 过多的波段不利于后续与多波段比值融合处理, 因此需进行主成分分析并选取合适的主分量。此方法主要是在 ENVI 5.3 软件中使用 Forward PCA Rotation New Statistics and Rotate 工具进行实现。

本研究将 Sentinel-2 影像的 13 个波段进行主成分分析, 得到各个主分量所占的特征值百分比。由表 2 可知, 第一主分量的特征值和累计特征值占比 76.93%; 第二主分量的特征值占比 14.10%, 其累计特征值占比 91.03%; 第三主分量的特征值占比 5.71%, 其累计特征值占比 96.74%。综上所述, 前三个主分量几乎涵盖了所有信息。为此, 我们将前三个主分量与 *NDVI*、*FAI* 进行融合, 得到最后用于绿潮信息提取的遥感影像。

表 2 主成分分析结果

Tab.2 Results of principal component analysis

PC	特征值	特征值百分比/(%)	累计特征值百分比/(%)
1	9.97E+05	76.93	76.93
2	1.83E+05	14.10	91.03
3	7.40E+04	5.71	96.74
4	2.00E+04	1.54	98.28
5	6.69E+03	0.52	98.80
6	5.54E+03	0.42	99.22
7	3.50E+03	0.27	99.49
8	2.57E+03	0.20	99.69
9	1.29E+03	0.10	99.79
10	8.88E+02	0.07	99.86
11	8.77E+02	0.07	99.93
12	6.30E+02	0.05	99.98
13	3.01E+02	0.02	100.00

### 1.3.4 分类方法

#### (1) 最大似然法

最大似然法通过假定各类分布函数为正态分布, 并选择训练区, 计算给定像元的归属概率, 像元最终被归到似然度最大的一类中<sup>[22]</sup>。由于 MLC 在一开始就假定各类分布函数为正态分布, 但实际分类过程与假定有一定差距, 因此 MLC 对训练样本的构建要求较高。此方法主要是在 ENVI 5.3 软件中使用 Maximum Likelihood Classification 工具进行实现。

#### (2) 随机森林

随机森林是一种以决策树为基分类器, 通过 Bagging 方法将多个决策树集成为随机森林的新型机器学习算法<sup>[23]</sup>, 相比于单棵决策树, 随机森林具有较强的挖掘能力以及较强的鲁棒性<sup>[24]</sup>, 不易陷入过拟合<sup>[25]</sup>。目前 RF 多用于处理高维数据, 被广泛应用于数学、地理、生物、生态、经济等领域<sup>[26]</sup>。此方法主要是在 ENVI 5.3 软件中使用 Random Forest Classification 工具进行实现。

#### (3) 支持向量机

支持向量机是一种采用最优分离超平面 (或超曲面) 来分离正负样本的二分类分类器。通常情况下, 使用 SVM 对中小型数据集进行地物分类时, 能取得较好的分类效果。但对于大型数据集以及多分类问题, SVM 训练所需时间较长<sup>[27]</sup>。此方法主要是在 ENVI 5.3 软件中使用 Support Vector Machine Classification 工具进行实现。

### 1.3.5 分类方法流程

绿潮信息提取主要流程包括: (1) 在 Sen2cor、SNAP 和 ENVI 软件中对获取到的 Sentinel-2 影像进行预处理, 包括辐射定标、大气校正、格式转换和镶嵌裁剪等工作。为了保证遥感数据各波段空间分辨率的一致性, 在 SNAP 平台中使用最邻近内插法将遥感数据各波段重新采样至 10 m。(2) 选取 B4/B8/B10 作为最佳波段组合, 通过多波段比值法计算 Sentinel-2 影像的 *NDVI* 和 *FAI*。(3) 由于 Sentinel-2 影像有 13 个波段, 为提高算法效率, 通过主成分分析选取合适的主分量, 得到主分量影像。(4) 利用 ENVI 5.3 软件 Layer Stacking 工具将步骤 3 得到的主

分量影像与 *NDVI*、*FAI* 进行融合, 获得融合后的影像数据。(5) 基于步骤 3 和步骤 4 分别得到的影像数据, 使用 MLC、RF 和 SVM 三种模型, 对绿潮信息进行提取, 并验证其精度。技术流程详见图 2。

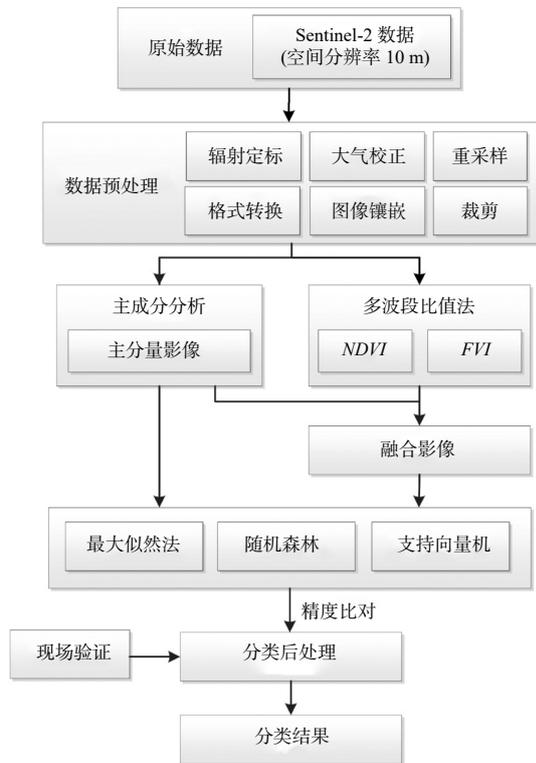


图 2 流程图

Fig. 2 Flow chart

## 2 结果与讨论

### 2.1 结果分析

#### 2.1.1 精度评价

本研究选择融合多波段比值法的 RF 模型与传统 MLC、RF 和 SVM 以及融合多波段比值法的 MLC 和 SVM 进行对照实验。表 3 为 6 种绿潮信息提取结果精度对比。MLC 和 SVM 为目前中低分辨率卫星影像研究中常用的分类算法。针对具体的分类类别, 传统 MLC 对绿潮的制图精度为 96.60%, RF 对绿潮的制图精度为 87.13%, SVM 对绿潮的制图精度为 93.09%。融合多波段比值法的 MLC、RF 和 SVM 三种分类算法在绿潮的制图精度上大幅度提高, 相较于前面三种分类算法分别提高了 2.84%、9.66% 和 1.84%。传统 MLC 对绿潮的用户精度为 75.30%, RF 对绿潮的用户精度为 100%, SVM 对绿潮的用户精度为 99.44%。融合多波段比值法的 MLC、RF 和 SVM 三种分类算法在绿潮的用户精度上有所降低, 相较于前面三种分类算法分别降低了 0.28%、0.11% 和 0.32%。

在海面非绿潮信息提取方面, 传统 MLC 的制图精度为 99.13%, RF 的制图精度为 100%, SVM 的制图精度为 99.99%。融合多波段比值法的 MLC 和 SVM 两种分类算法在非绿潮的制图

表 3 不同分类算法精度

Tab.3 Accuracy of different classification algorithms

方法	分类类别	制图精度/(%)	用户精度/(%)	总体精度/(%)	Kappa系数
MLC	绿潮	96.60	75.30	99.06	0.84
	非绿潮	99.13	99.91		
RF	绿潮	87.13	100.00	99.66	0.93
	非绿潮	100.00	99.65		
SVM	绿潮	93.09	99.44	99.80	0.96
	非绿潮	99.99	99.81		
融合多波段比值法的MLC	绿潮	99.44	75.02	99.10	0.85
	非绿潮	99.09	99.98		
融合多波段比值法的RF	绿潮	96.79	99.89	99.92	0.98
	非绿潮	100.00	99.92		
融合多波段比值法的SVM	绿潮	94.96	99.12	99.84	0.97
	非绿潮	99.98	99.86		

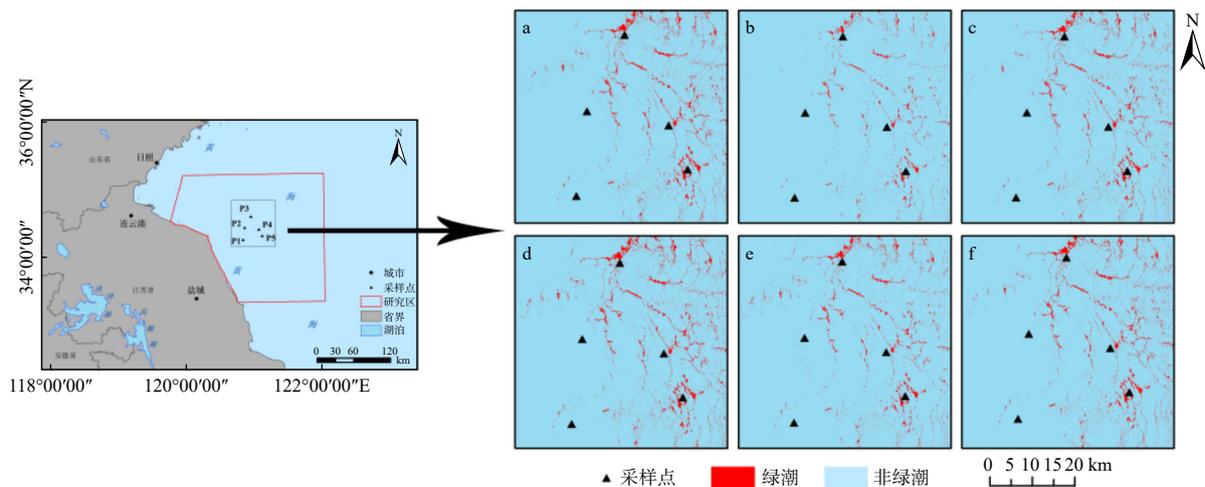
精度上略微降低,相较于两种传统分类算法分别降低了 0.04% 和 0.01%,而融合多波段比值法的 RF 算法未有任何变化。传统 MLC 对非绿潮的用户精度为 99.91%,RF 的用户精度为 99.65%,SVM 的用户精度为 99.81%。融合多波段比值法的 MLC、RF 和 SVM 三种分类算法在非绿潮的用户精度上有所提高,相较于前面三种分类算法分别提高了 0.07%、0.27% 和 0.05%。

从总体精度来看,传统的 MLC 总体精度为 99.06%,Kappa 系数为 0.84,RF 总体精度为 99.66%,Kappa 系数为 0.93,SVM 总体精度为 99.80%,Kappa 系数为 0.96。融合多波段比值法后,MLC 总体精度为 99.10%,Kappa 系数为 0.85,RF 总体精度为 99.92%,Kappa 系数为 0.98,SVM 总体精度为 99.84%,Kappa 系数为 0.97。

### 2.1.2 现场验证

本研究选用 2021 年 6 月 6—8 日在南黄海海域现场同步调查的数据,用以辅助训练样本的选择以及六种绿潮信息提取结果的可靠性验

证。受船舶航程和海面天气影响,我们选取了 5 处典型绿潮分布跟踪采样点(图 3 左和图 4),并加载六种提取结果(图 3 右)。从图 3、图 4 和表 4 中可发现,绿潮漂浮在海水表面,利用 Sentinel-2 影像基本可在广阔海域中将其与海水区分。1 号跟踪点存在不均匀分布、零星漂浮的点状绿藻,最大面积不超过  $0.3 \text{ m}^2$ ,Sentinel-2 影像中未探测到此处漂浮绿潮信息。2 号跟踪点已形成条状绿潮带,但分布较为稀疏,密集度不足 15%,影像可观测到微弱漂浮绿藻信息,同时在六种绿潮信息提取结果中发现仅融合多波段比值法的 RF 模型可在 P2 点提取到少量有效的绿潮信息(图 3e)。3 号、4 号和 5 号跟踪点均存在大面积条带状分布的绿藻,影像中可以清晰探测到,六种分类模型均可提取到绿潮信息(图 3),但在绿潮边缘局部区域有所差异。综上,本研究中六种分类模型提取到绿潮信息基本能够与现场数据相匹配,其中融合多波段比值法的 RF 模型对海面少量漂浮绿潮的信息提取更为准确。



注: a.传统 MLC; b.传统 RF; c.传统 SVM; d.融合多波段比值法的 MLC; e.融合多波段比值法的 RF; f.融合多波段比值法的 SVM

图 3 2021 年 6 月 6—8 日现场验证空间分布及绿潮信息提取结果局部空间分布

Fig. 3 Spatial distribution of on-site verification and local spatial distribution of green tide information extraction results on June 6—8, 2021

### 2.1.3 分类结果

为了进一步研究不同机器学习算法在绿潮与非绿潮之间分类性能的差异,统计了六种分类算法提取的各类地物面积。传统 MLC、传统 RF、传统 SVM、融合多波段比值法的 MLC、融合多波段比值法的 RF 和融合多波段比值法的

SVM 等六种方法提取绿潮的空间分布见图 5。六种分类模型提取绿潮结果的空间分布呈“北密南疏”,近岸海域基本无绿潮分布。MLC 提取结果有个别区域明显异于其他两种模型,而 SVM 模型在研究区东北区识别绿潮信息多于 RF 模型。而融合多波段比值法后三种分类模型

表 4 2021 年 6 月 6—8 日不同站位 Sentinel-2 影像中漂浮绿藻信号的现场验证

Tab.4 Field verification of floating green algae signal by Sentinel-2 image at different stations on June 6—8, 2021

站位	经度/° N	纬度/° E	时间	分布状态
P1	120.866	34.287	09:06	无
P2	120.877	34.458	10:10	稀疏分布
P3	120.938	34.599	11:25	宽条带状
P4	121.056	34.409	14:06	宽条带状
P5	121.106	34.316	15:33	宽条带状

对于较为稀疏的绿潮识别能力均有所提高。另外,我们计算发现传统 MLC 提取绿潮的面积为 792.78 km<sup>2</sup>(图 5 a), 占研究区总面积的 1.82%; 传统 RF 提取面积为 246.57 km<sup>2</sup>(图 5 b), 占比为 0.57%; 传统 SVM 提取面积为 597.19 km<sup>2</sup>(图 5 c), 占比为 1.37%。融合多波段比值法的 MLC、RF 和 SVM 三种方法提取的绿潮覆盖率相比传统的三种方法均有所提高, 其中融合多波段比值法的 MLC 提取的绿潮覆盖率提高了 0.14%(图 5 d),

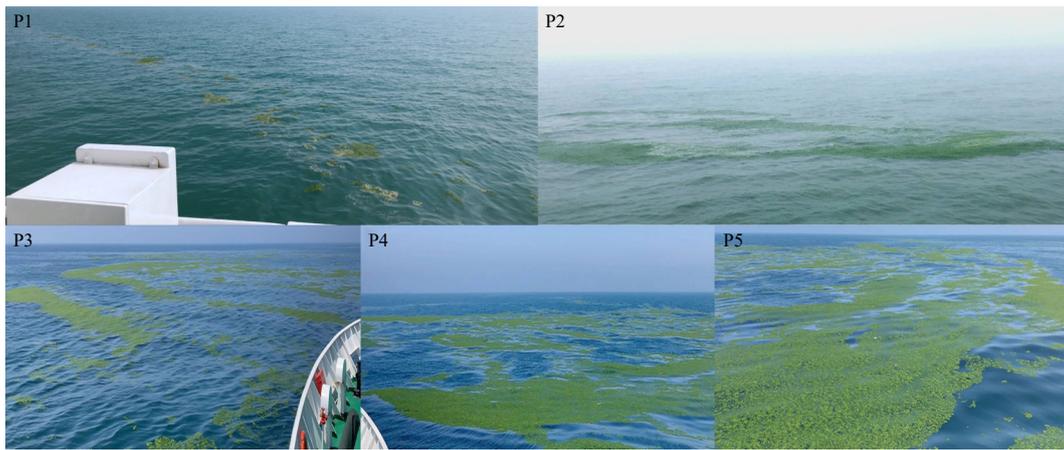
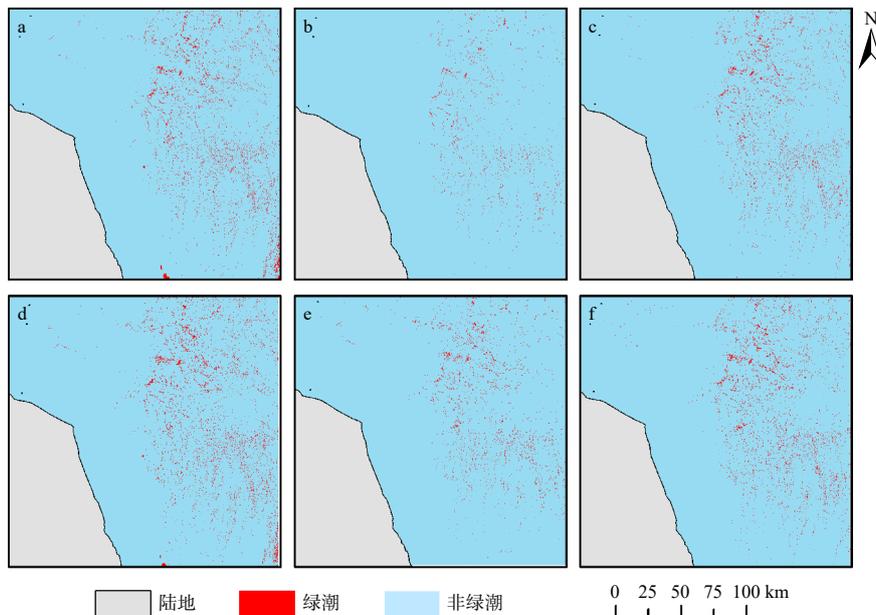


图 4 2021 年 6 月 6—8 日 Sentinel-2 影像中漂浮绿藻信号的现场验证

Fig. 4 Field verification of floating green algae signal from Sentinel-2 images on June 6—8, 2021



注: a.传统 MLC; b.传统 RF; c.传统 SVM; d.融合多波段比值法的 MLC; e.融合多波段比值法的 RF; f.融合多波段比值法的 SVM

图 5 绿潮提取空间分布

Fig. 5 Spatial distribution of green tide extraction

融合多波段比值法的 RF 提高了 0.38%(图 5 e), 融合多波段比值法的 SVM 提高了 0.16%(图 5 f)。

另外,我们在研究区中选取了四处绿潮典型区域(密集区、稀疏区、线型和散型)进行对比分析(图 6)。研究发现在绿潮密集区,六种分类模型均可较为准确地识别绿潮的存在,在传统的三种模型中,MLC 和 SVM 表现明显优于 RF,但是 MLC 和 SVM 均存在错分现象,融合多波段指数之后,MLC 和 SVM 分类结果无明显变化,而 RF 分类模型的漏分现象极大改善,与原始影像

基本一致,明显优于其他五种。在稀疏区,传统三种模型对绿潮的漏分现象较为严重,其中传统 RF 模型甚至无法识别出绿潮,融合多波段指数之后的三种模型均有所改善,其中融合 RF 模型表现最好。在线型绿潮中,传统 MLC 明显优于其他两种模型,融合多波段指数之后 MLC 和 RF 表现基本一致,优于 SVM 模型。对于散型绿潮三种模型表现较差,均出现对绿潮的错分,其中 RF 模型略优于另外两种模型,而融合多波段指数之后绿潮的错分现象并无明显改善。

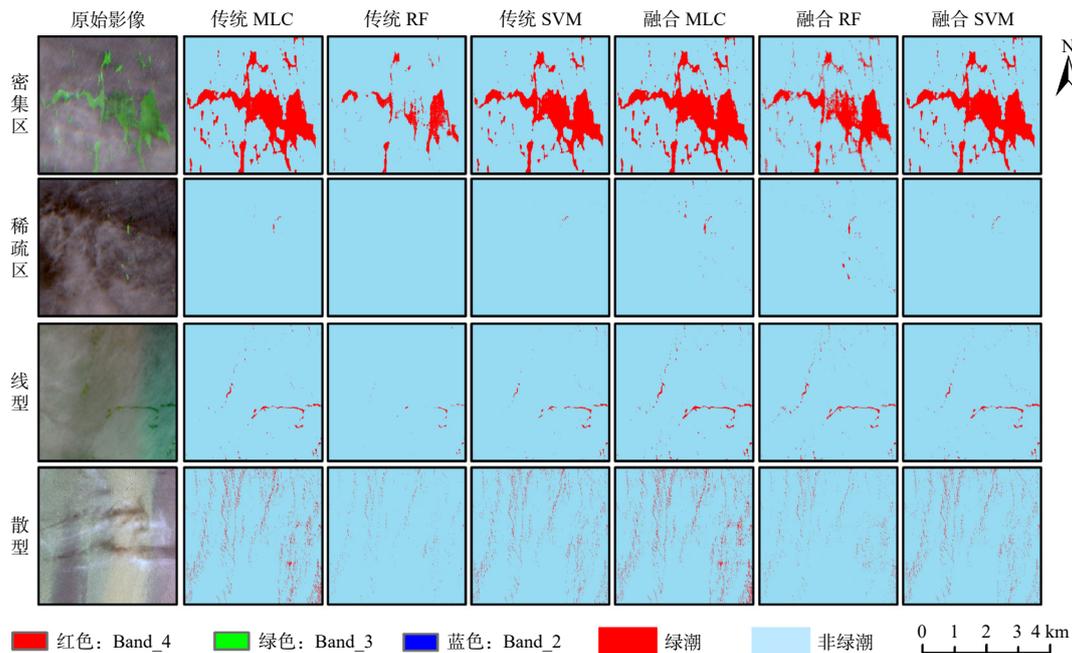


图 6 四种典型空间区域的绿潮提取空间分布

Fig. 6 Spatial distribution of green tide extraction in four typical spatial regions

## 2.2 讨论

### 2.2.1 分类模型的对比

本研究将传统 MLC、RF 和 SVM 以及融合多波段比值法的 MLC、RF 和 SVM 等六种模型的结果进行对比。在分类结果精度评价中,制图精度用以表征模型对绿潮的漏分现象,用户精度用以表征模型对绿潮的错分现象。在传统的三种分类模型中,MLC 模型对绿潮的错分现象较严重,RF 模型对绿潮的漏分现象较严重,而 SVM 模型的制图精度和用户精度均未出现过低的情况。对比融合多波段比值法后的三种分类模型,我们发现融合多波段比值法可以有效改善

对绿潮的漏分现象,进而提高绿潮的制图精度,与梁锦涛等<sup>[16]</sup>研究发现一致(融合指数可提高分类模型对相应地物的准确性)。造成这一现象的原因可能是 *NDVI* 和 *FAI* 两种指数能突出影像中藻类植物与其他地物的差异,从而极大提高了对绿潮信息的提取。传统三种分类模型中,只有传统 RF 模型中绿潮的制图精度最低,因此,融合多波段比值法对 RF 模型的精度提升最大,而对另外两种模型的提升较小。

本研究仅选择了两种地物,在精度评价中可能会弱化融合多波段比值法的 RF 模型相较于其他五种分类模型的优势,故我们又基于真彩色合

成影像(图 7)对六种模型分类结果的空间分布进行整体及局部的对比分析。整体来看, SVM 模型、融合多波段比值法的 RF 模型和 SVM 模型明显优于 MLC 模型。基于 MLC 的两种分类模型在提取绿潮信息过程中,均会将研究区南部的其他藻类植物错分为绿潮(图 5 a, 图 5 d)。其原因可能是 *NDVI* 指数和 *FAI* 指数难以区分不同藻类之间的差异,融合这两种指数对该现象并无明显改善。而在选取的四处绿潮典型区域(密集区、稀疏区、线型和散型)中,六种分类模型的表现与精度评价基本一致,融合多波段比值法的 RF 模型明显优于基于 MLC 的两种分类模型及传统 RF 模型,略优于基于 SVM 的两种分类模型。另外, Sentinel-2 影像空间分辨率较高,我们在提取过程中发现 SVM 模型运算耗时过长,与贺丹等<sup>[27]</sup>结论一致。因此,基于高分辨率遥感影像提取绿潮信息时,融合多波段比值法的 RF 模型更为实用。

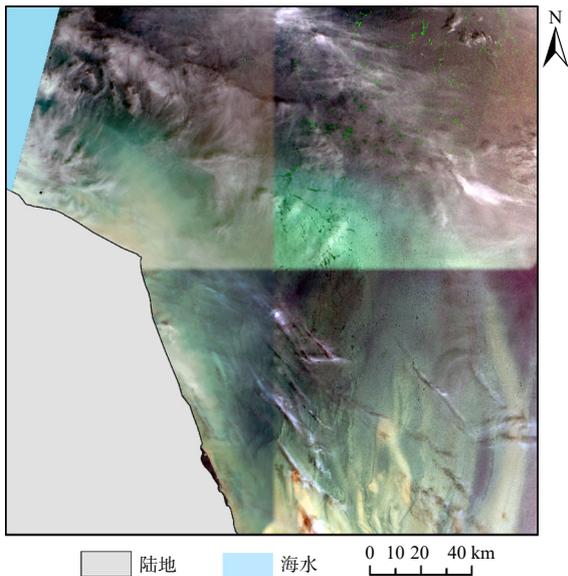


图 7 研究区 Sentinel-2 真彩色合成影像(R: B4, G: B3, B: B2)

Fig. 7 Sentinel-2 natural color composite image of the study area (R: B4, G: B3, B: B2)

### 2.2.2 不足与展望

在对六种模型分类结果进行典型区域局部对比时,我们发现六种模型对于散型绿潮的提取效果较差。基于 RF 模型的两种分类模型虽优于另外四种,但仍出现大量绿潮错分现象。这种情

况可能是由于散型绿潮的像元较小,同时分布较为密集,分类模型将散型绿潮周围的影像噪点错分为绿潮。因此在提取散型绿潮时,应考虑使用其他地物分类方法。

黄海绿潮发生初期斑块面积较小,中低分辨率光学数据难以对其进行识别并可能造成较大误差。但中低分辨率光学数据常用的分类算法在高分辨率光学数据中并不能得到较好效果,而本研究提出的融合多波段法和 RF 的分类方法大幅提高了绿潮信息提取效果,充分发挥了高分辨率光学影像的优势。未来研究中,我们将尝试融合更多的指数,来提高分类模型对黄海绿潮的提取精度。另外,我们还将采用多种技术手段(如无人机)获取充足的现场调查数据及更高空间分辨率的影像数据,用以验证黄海绿潮提取方法的可靠性。

### 3 结论

(1)传统 MLC、RF 和 SVM 以及融合多波段比值法的 MLC、RF 和 SVM 等六种模型均可对绿潮进行有效提取。其中传统的三种模型中 SVM 分类精度最高,MLC 对绿潮错分较为严重,RF 对绿潮漏分较为严重;融合多波段比值法后 RF 分类精度最高, SVM 次之,MLC 对绿潮的错分现象无明显改善。总体上,融合多波段比值法的 RF 分类模型对苏北海域绿潮信息的提取精度最高,提取结果最准确, Kappa 系数为 0.98, 总体精度高达 99.92%。

(2)*NDVI* 和 *FAI* 能够极大突出遥感影像中藻类植物和其他地物的差异,有效改善分类模型对绿潮的漏分现象,以提高信息提取的准确性。融合水体植被指数的 RF 方法在绿潮错分方面明显优于传统 MLC、RF 和 SVM 以及融合多波段比值法的 MLC 和 SVM 等五种模型。

### 参考文献:

- [1] 于仁成, 孙松, 颜天, 等. 黄海绿潮研究: 回顾与展望[J]. *海洋与湖沼*, 2018, 49(5): 942-949.
- [2] SMETACEK V, ZINGONE A. Green and golden seaweed tides on the rise[J]. *Nature*, 2013, 504(7478): 84-88.
- [3] 蒋雪蕾, 周晓见, 林佳宁, 等. 黄海洋苔绿潮生态效应研究进展[J]. *海洋环境科学*, 2021, 40(4): 647-652.

- [4] 王宗灵, 傅明珠, 肖洁, 等. 黄海浒苔绿潮研究进展[J]. 海洋学报, 2018, 40(2): 1-13.
- [5] 王国伟, 李继龙, 杨文波, 等. 利用MODIS和RADARSAT数据对浒苔的监测研究[J]. 海洋湖沼通报, 2010 (4): 1-8.
- [6] 邢前国, 郑向阳, 施平, 等. 基于多源、多时相遥感影像的黄河、东海绿潮影响区检测[J]. 光谱学与光谱分析, 2011, 31(6): 1644-1647.
- [7] SON Y B, MIN J E, RYU J H. Detecting massive green algae (*Ulva prolifera*) blooms in the Yellow Sea and East China Sea using Geostationary Ocean Color Imager (GOCI) data[J]. *Ocean Science Journal*, 2012, 47(3): 359-375.
- [8] 袁超, 张靖宇, 肖洁, 等. 基于哨兵2号卫星遥感影像的2018年苏北浅滩漂浮绿藻时空分布特征研究[J]. 海洋学报, 2020, 42(8): 12-20.
- [9] HU C M, LI D Q, CHEN C S, et al. On the recurrent *Ulva prolifera* blooms in the Yellow Sea and East China Sea[J]. *Journal of Geophysical Research:Oceans*, 2010, 115(C5): C05017.
- [10] NARIN O G, BAYIK C, ABDIKAN S, et al. Using RVI and NDVI time series for cropland mapping with time-weighted dynamic time warping[C]//Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Castelo Branco: ISPRS, 2022.
- [11] 李敬虎, 邢前国, 郑向阳, 等. 基于深度学习的无人机影像夜光藻赤潮提取方法[J]. 计算机应用, 2022, 42(9): 2969-2974.
- [12] OIRY S, BARILLÉ L. Using sentinel-2 satellite imagery to develop microphytobenthos-based water quality indices in estuaries[J]. *Ecological Indicators*, 2021, 121: 107184.
- [13] 张佩莹, 张方方, 李俊生, 等. 基于Sentinel-2影像特征优化的于桥水库水生植被提取[J]. 生态科学, 2023, 42(1): 40-48.
- [14] 蔡晓晴. 基于静止轨道海洋水色卫星数据的绿潮遥感探测方法和逐时变化特征研究[D]. 青岛: 中国海洋大学, 2014.
- [15] 朱青, 国佳欣, 郭熙, 等. 基于随机森林算法的土壤侵蚀影响因子研究——以赣江上游流域为例[J]. 水土保持通报, 2020, 40(2): 59-68.
- [16] 梁锦涛, 陈超, 张自力, 等. 一种融合指数与主成分分量的随机森林遥感图像分类方法[J/OL]. 自然资源遥感: 1-8[2023-08-29]. <http://kns.cnki.net/kcms/detail/10.1759.P.20230619.1004.010.html>.
- [17] 中华人民共和国自然资源部. 2021年中国海洋灾害公报[R/OL]. 北京: 中华人民共和国自然资源部, 2022. [2023-07-29]. [http://gi.mnr.gov.cn/202205/t20220507\\_2735508.html](http://gi.mnr.gov.cn/202205/t20220507_2735508.html).
- [18] 刘材材, 徐韧, 何培民, 等. 南黄海绿潮暴发与紫菜养殖的关系[J]. 海洋科学, 2017, 41(2): 35-43.
- [19] 王怡人, 王胜强, 喻樾, 等. 一种提取南黄海浒苔的自适应阈值遥感算法[J]. 遥感信息, 2021, 36(2): 120-129.
- [20] HU C M. A novel ocean color index to detect floating algae in the global oceans[J]. *Remote Sensing of Environment*, 2009, 113(10): 2118-2129.
- [21] KOUTSIAS N, MALLINIS G, KARTERIS M. A forward/backward principal component analysis of Landsat-7 ETM+ data to enhance the spectral signal of burnt surfaces[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2009, 64(1): 37-46.
- [22] 邓书斌, 陈秋锦, 杜会建. ENVI遥感图像处理方法[M]. 2版. 北京: 高等教育出版社, 2014: 477.
- [23] DE'ATH G, FABRICIUS K E. Classification and regression trees: a powerful yet simple technique for ecological data analysis[J]. *Ecology*, 2000, 81(11): 3178-3192.
- [24] 吴兴惠, 周玉萍, 邢海花. 集成学习之随机森林分类算法的研究与应用[J]. 电脑知识与技术, 2020, 16(21): 26-27.
- [25] JI X, YANG B S, TANG Q H. Seabed sediment classification using multibeam backscatter data based on the selecting optimal random forest model[J]. *Applied Acoustics*, 2020, 167: 107387.
- [26] RODRIGUEZ-GALIANO V F, GHIMIRE B, ROGAN J, et al. An assessment of the effectiveness of a random forest classifier for land-cover classification[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2012, 67: 93-104.
- [27] 贺丹, 张静. 深度学习方法在松山湖地区地物分类中的应用[J]. 科技创新导报, 2020, 17(33): 21-25.

( 本文编辑: 曲丽梅 )