

引用格式: ZHU Xiufang, LI Yuan, GUO Rui. Rapid detection and identification of water anomalies based on Isolated Forest[J]. Remote Sensing Technology and Application, 2023, 38(5): 1126-1135. [朱秀芳, 李原, 郭锐. 基于孤立森林的水体异常快速发现与识别[J]. 遥感技术与应用, 2023, 38(5): 1126-1135.]  
DOI: 10.11873/j.issn.1004-0323.2023.5.1126

## 基于孤立森林的水体异常快速发现与识别

朱秀芳<sup>1,2</sup>, 李原<sup>3,4</sup>, 郭锐<sup>1,2</sup>

1. 北京师范大学遥感科学国家重点实验室, 北京 100875;
2. 北京师范大学地理科学学部遥感科学与工程研究院, 北京 100875;
3. 内蒙古大学生态与环境学院, 内蒙古呼和浩特 010021;
4. 内蒙古自治区河流与湖泊生态重点实验室, 内蒙古呼和浩特 010021)

**摘要:** 现有异常水体的检测研究通常针对特定区域、特定数据源和特定时相开展, 且往往是事后的检测而非实时的监测, 不能很好地服务于水体异常快速发现与识别的业务需求。为此, 提出了一种基于无监督的孤立森林加决策规则(U-IForest-SD)的水体异常信息提取方法, 并以Landsat与Sentinel的数据作为测试数据, 以青岛浒苔、松雅湖黑臭水、墨西哥湾溢油为案例, 对比了支持向量机、监督孤立森林以及U-IForest-SD 3种方法识别浒苔、黑臭水和溢油的精度。研究结果显示: 该方法对于3种异常类型的总体识别精度都在90%以上, Kappa系数都在0.8以上, 整体精度高于监督孤立森林但略低于SVM。该算法只需要输入单期影像, 无需训练样本, 具有可移植性好、普适性强、自动化程度高的优点。此外, 该方法可以有效地避免“假警”和“虚警”的发生, 在异常水体的快速发现和识别业务中有很好的应用前景。

**关键词:** 异常检测; 溢油; 黑臭水; 浒苔

**中图分类号:** X87; TP79 **文献标志码:** A **文章编号:** 1004-0323(2023)05-1126-10

### 1 引言

水生环境是生态系统的重要组成部分, 近年来随着人类活动的加剧, 近海与内陆的水环境中水污染事件的爆发频率显著增加。国内外很多学者使用遥感技术进行水环境的研究<sup>[1-3]</sup>, 包括对于赤潮<sup>[4-7]</sup>、浒苔<sup>[8-10]</sup>、黑臭水<sup>[11-12]</sup>、溢油<sup>[13-15]</sup>等水体污染事件的监测。然而现有研究通常针对特定区域、特定数据源和特定时相开展, 方法缺乏普适性。

人为或自然因素造成的水体污染事件均可以视为水体异常事件, 可以通过遥感异常检测技术来进行识别。遥感图像异常目标检测技术发展至今, 已有大量的算法被提出, 如基于深度学习模型<sup>[16]</sup>, 基于支持向量<sup>[17]</sup>、基于自适应阈值<sup>[18]</sup>、基于子空

间<sup>[19]</sup>、基于聚类<sup>[20]</sup>等的异常检验算法。然而, 遥感图像数据量大、维度高且冗余现象严重, 常见的大部分算法都有着较高的计算复杂度, 往往理论上性能优越, 但效率上难以满足实时遥感服务的需要。因此, 遥感图像异常目标检测技术发展至今, 如何实现高精度、高效率异常目标检测任务, 一直是研究者们努力的方向<sup>[21]</sup>。

孤立森林是一种适用于异常检测的算法<sup>[22]</sup>。该算法不涉及诸如求导、求图像的协方差矩阵等复杂高阶运算, 可以应用到数据量较大的异常检测问题中<sup>[23]</sup>。然而, 目前孤立森林在遥感图像异常目标检测上的相关研究还很少, 且大多集中在陆地高光谱图像上<sup>[24]</sup>, 对于水体异常检测的研究还没有报道。

基于上述考虑, 本研究拟发展一种孤立森林与

收稿日期: 2022-09-12; 修订日期: 2023-08-05

基金项目: 国家自然科学基金重大项目(41292583)。

作者简介: 朱秀芳(1982-), 女, 浙江天台人, 教授, 主要从事遥感应用研究。E-mail: zhuxiufang@bnu.edu.cn

决策树相结合的服务于水体异常事件快速识别与诊断的算法,在保证算法尽量自动化、快速化的前提下可以得到满足精度要求的识别结果,以此来拓展孤立森林算法在遥感图像异常检测领域的迁移应用,为针对一些实际应用情形开发具备工程化前景且性能优秀的算法提供借鉴。

## 2 水体异常事件与数据收集

### 2.1 典型水体异常事件

水体异常事件种类繁多,包括有机氮磷污染、COD超标、重金属污染、藻类爆发、溢油等。其中部分异常并不会明显改变水体的光谱信息,使用遥感手段难以进行有效的监测。综合考虑水体异常事件的特征和遥感监测的基本原理,本研究将基于遥感的水体异常事件监测重点放在了藻类爆发、黑臭水体以及溢油3种突发性灾害上。这3种异常事件在近些年对内陆或近海水体均造成过严重的危害,如发生在黄海沿岸地区的浒苔<sup>[25]</sup>、湖南省长沙市大范围黑臭水体污染<sup>[26]</sup>、墨西哥湾的溢油事件<sup>[27]</sup>。同时这3种类型的水体异常都不同程度的改变了水体的光谱反射曲线,为其遥感监测提供了可行的基础。为了进行算法测试,针对上述3类水体异常,分别选择了一个典型的案例进行研究。具体信息如下:

#### (1) 青岛浒苔

水体典型异常事件中藻类等水生植物爆发性繁殖引发的异常事件包括水葫芦、赤潮、浒苔等。其中青岛浒苔在近近年来以爆发迅猛覆盖面积大而受到极大的关注。浒苔的大量繁殖已经给沿岸的生态环境以及渔业养殖业带来了重大危害。

#### (2) 松雅湖黑臭水体

黑臭水体是由于水环境遭受了超过其自净能力的有机污染。有机物的好氧分解使水体中耗氧

速率大于复氧速率,造成水体缺氧,致使有机物降解不完全生成硫化氢等发臭物质同时产生黑色物质,使水体发黑发臭。根据全国地级及以上城市黑臭水体名单<sup>[28]</sup>,长沙市主要河流湖泊在2013年到2018年之间存在大面积黑臭现象。进一步参考陈帅<sup>[26]</sup>的研究,收集对应的遥感数据进行案例研究。

#### (3) 墨西哥湾溢油

溢油事故的发生会对水生态环境及与水有关的经济产业造成极其恶劣的影响。世界范围内每年都有大量的石油及其相关产品侵入海洋、陆地及地下水领域。从NOAA给出的每年石油和化学品泄漏事件数据集来看溢油事件时有发生,墨西哥湾作为美国重要的石油产地<sup>[29-30]</sup>,是溢油事件爆发的重灾区。依据NOAA给出的溢油事件爆发点数据,结合海表面油膜的纹理及光谱特性<sup>[31-32]</sup>,筛选近年来出现的溢油事件,最终选定2020年5月的新奥尔良东部近海地区数据作为测试数据集。

### 2.2 数据与预处理

本研究以Landsat与Sentinel数据作为输入数据集(图1和表1)。Landsat系列卫星发展时间长且可免费下载,使用Landsat数据进行灾害检测的案例较多,为数据与案例的寻找提供了一定的便利。另外,Landsat数据具有较好的空间分辨率(例如Landsat 8的全色波段分辨率为15 m,其余波段的分辨率为30 m)、光谱跨度较大(从可见光至中红外),与目前大多数在轨卫星的光谱设置均有交集。Sentinel数据的发展时间虽然没有Landsat系列长,但最高10 m的空间分辨率以及在可见光至中红外谱段更加精细的划分,使其近年来成为能够免费获取的数据中最受欢迎的数据之一。Sentinel还提供雷达数据,且多星的设置使其在月内能够对目标地进行多次重访,可以满足许多特定工作的需求。

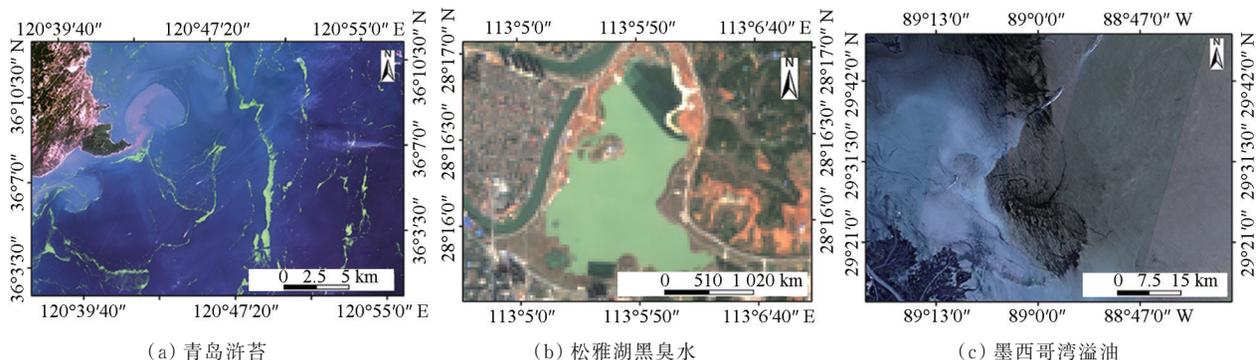


图1 研究区概况

Fig.1 Overview of the study area

表1 数据详情  
Table 1 Data details

数据类型	获取时间	数据描述	用途
Landsat 8	2021-07-09	青岛浒苔发生时的数据	测试算法提取浒苔的适用性
Landsat 8	2021-09-11	青岛浒苔结束后的数据	测试算法在正常水体中是否会错误识别异常
Landsat 8	2016-11-28	松雅湖黑臭水体发生时的数据	测试算法提取黑臭水体的适用性
Sentinel-2	2020-05-02	墨西哥湾溢油发生时的数据	测试算法提取溢油的适用性

本研究对 Landsat 数据使用 ENVI 的 FLAASH 模块、对 Sentinel 数据使用 Sen2Cor 插件进行大气校正,并通过重采样使 Landsat 8 的各波段的空间分辨率都保持为 30 m、Sentinel-2 的各波段的空间分辨率都保持为 10 m。通过目视解译数字化水体边界,以提取出水体范围。此外,根据已有文献<sup>[26,31-32]</sup>和经验,目视确定了浒苔像元 260 个、黑臭水像元 132 个、溢油像元 213 个,对应分别在 3 个研究区选择了正常水体像元 372、274 和 223 个,合计 1 474 个样本,用这些检验样本验证不同方法识别出异常水体像元的精度。对于监督分类方法,还需要训练样本,分别针对不同的数据源(Landsat 8 和 Sentinel-2)目视确定训练样本,各数据源上的样本个数为 180 个,其中 30% 为异常水体像元,70% 为正常水体像元。

### 3 研究方法

#### 3.1 方法介绍

研究提出一种基于无监督的孤立森林(U-IForest)和决策规则的水体异常信息自动化提取方法。其中无监督的孤立森林用来快速发现异常像元,决策规则用来诊断异常像元的异常类型。

孤立森林算法(IForest)是由 Liu 等<sup>[34]</sup>在第八届 IEEE 数据挖掘国际会议上首次提出,主要用于数据挖掘中的离群点检测。孤立的意思是“将具有一定特性的样本从其他样本数据中分离出来”。这里的特性有两层含义:①样本在整个数据中的占比非常少;②样本具有与周围样本数据明显不同的值。对于遥感影像中的异常目标检测问题,孤立的概念同样适用,这是由于异常像元往往具有“数量少且不同”的特性,因此更容易被孤立。当影像中的某些像元在由若干个孤立树组成的孤立森林中均具有较短的路径长度时,这些像元就是异常的。对图像进行无监督孤立森林操作后,会得到每个像元的异常程度值,通过设定异常程度的阈值,进一步将异常程度图转换为-1、1 二值图(即异常非异常二值图)。因此,阈值的选择是关键。为了提高算法的

普适性和自动化程度,使用统计学对异常值的定义,利用标准差法进行阈值确定,具体是均值±标准差的方法,即如果某个水体像元的异常程度的取值在异常程度均值±标准差的范围之外则被认为是异常的水体像元。

在得到异常像元后,还需要进一步判断异常的类型,图 2 显示了 3 类异常水体的 NDVI 值域。图 2 (b)为 3 种异常水体和正常水体的归一化植被指数(NDVI)箱线图。由图可以看出,浒苔在近红外波段表现出高反射率这一点与正常水体以及其他两种异常事件迥然不同,溢油的反射率整体低于正常水体和其他两种异常水体。3 种异常水体的 NDVI 差异明显,在确定异常像元的基础上,通过 NDVI 来识别异常的类型。表 2 总结了 3 种异常水体的 NDVI 的取值范围(均值±2 倍标准差),将待判定的异常水体像元的 NDVI 值与表 2 中 3 种异常的取值范围进行比对,落入哪个区间就算是哪种异常水体。在经过上述判断后,可能有些异常水体像元没有被分到任何一个异常类型中,也可能存在错分。我们假设对应一个监测区域,异常水体的异常类型只有一种。因此,进一步规定如果有 90% 的异常像元没有被分入任何一个异常类型,则将所有的异常像元重新判定为正常水体像元,即该区域没有发生水体异常事件,否则将总数占比最大的异常类型作为该图像上异常水体像元的最终异常类型,把判读为其他异常类型的像元重编码至该类型。

#### 3.2 对比实验设计

为了说明所提出方法的有效性,本研究同时使用了 3 类算法进行异常水体信息提取:①基于传统监督分类方法(支持向量机)的异常水体信息提取;②基于监督孤立森林(S-IForest)的异常水体信息提取;③基于无监督的孤立森林(U-IForest)和决策规则的异常水体信息提取。此外,在无监督孤立森林中还进一步对比了不同离群值确定方法以及是否使用波段压缩方法对识别精度的影响。其中,离群值确定的方法使用了四分位距法和标准差法两种。四分位距法中将异常值定义为小于  $QL-n$

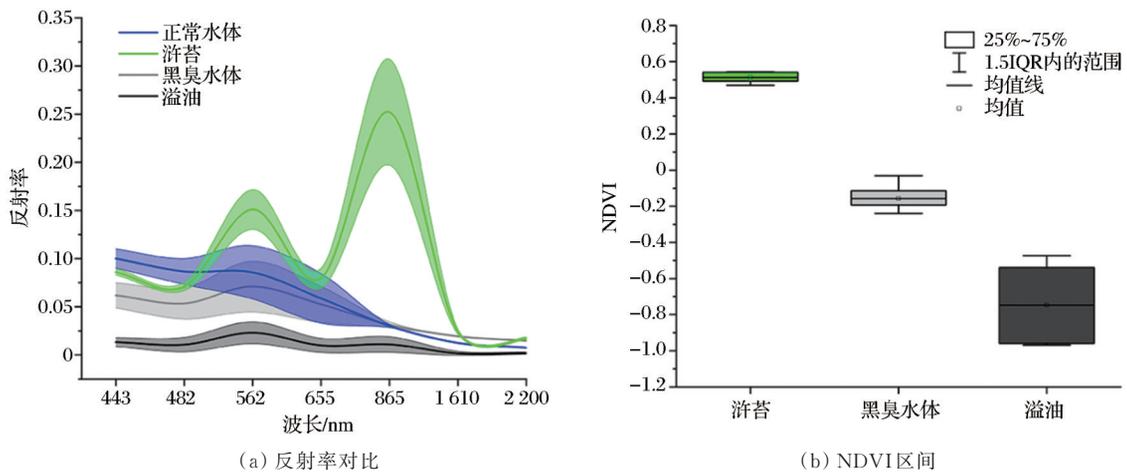


图 2 正常水体与异常水体的对比

Fig.2 Comparison between normal and abnormal water bodies

表 2 异常水体 NDVI 统计

Table 2 NDVI statistics of abnormal water bodies

类型	NDVI				
	最小值 Min	最大值 Max	均值 Mean	标准 差 $\sigma$	范围 (Mean $\pm$ 2 $\sigma$ )
浒苔	0.16	0.61	0.408	0.13	(0.15, 0.67)
黑臭水体	-0.303	-0.13	-0.204	0.049	(-0.302, -0.106)
溢油	-0.98	0	-0.88	0.23	(-1, -0.42)

倍 IQR 或大于  $QU+n$  倍 IQR 的值,  $n$  一般取 1.5, 实验测试了  $n=1$  和  $n=1.5$  两种情况。QL 为下四分位数, 表示全部观察值中有 25% 的数据取值比它小; QU 为上四分位数, 表示全部观察值中有 25% 的数据取值比它大; IQR 为四分位数间距, 是上四分位数 QU 与下四分位数 QL 之差, 其间包含了全部观察值的 50%。标准差法通过计算变量的均值和标准差来确定的, 标准差本身可以体现变量的离散程度。在离群值处理过程中, 可通过用均值  $\pm n$  倍标准差来衡量变量取值与平均值的距离,  $n$  一般取 2, 实验测试了  $n=1$  和  $n=2$  两种情况。因此, 在基于无监督的孤立森林(U-IForest)的异常水体信息提取中又具体包括了 4 种方法: U-IForest-IQR、U-IForest-1.5IQR、U-IForest-SD、U-IForest-2SD 分别对应 1 倍四分位距法、1.5 倍四分位距法、1 倍标准差法和 2 倍标准差法, 其中 U-IForest-SD 即为本文提出的方法。上述算法都对原始数据进行主成分分析, 选择前 3 个主成分进行处理的。为了对比使用和不使用波段压缩对本研究方法精度的影响, 我们也测试了利用原始所有波段数据执行 U-IForest-SD 算法来识别异常水体的效果, 将此方法记作 UP-CA-U-IForest-SD。

此外, 为了说明联合 U-IForest 和决策规则的必要性, 对比了 U-IForest-SD 和只用决策规则进行 3 类异常水体识别的效果。于此同时, 考虑以往研究往往只是基于有异常发生的水体的遥感影像进行测试研究, 此时是针对已经出现异常的水体开展的异常检测。然而, 在实时水体监测中, 大多数情况下我们所面对的是正常水体, 所发展的算法是否会虚报异常也是需要考虑的问题。为此, 在青岛浒苔案例的基础上, 选择浒苔结束后该区域的一景正常水体影像进行 U-IForest-SD 方法的测试。

### 3.3 实验流程

试验基本流程如图 3 所示。主要包括 6 个步骤: ①案例数据的准备和预处理; ②样本数据的准备; ③基于传统监督分类方法的异常识别; ④基于监督孤立森林(S-IForest)的异常水体信息提取; ⑤基于无监督的孤立森林(U-IForest)和决策规则的异常水体信息提取, 具体包括 U-IForest-IQR、U-IForest-1.5IQR、U-IForest-SD、U-IForest-2SD; ⑥利用混淆矩阵进行不同检测方法的精度评价和对比。其中, 预处理主要包括大气校正和基于主成分分析的波段压缩, 压缩后选取前 3 个波段参与后续处理。样本数据的准备包括两个部分: 服务于步骤 3 和步骤 4 的训练样本数据的制作; 服务于步骤 6 验证样本数据的制作。训练和验证样本数据均结合前人研究成果, 通过目视解译得到, 涵盖 3 类异常水体和正常水体。精度验证阶段, 使用验证数据集构建混淆矩阵, 得到各类异常水体像元识别的用户精度(User Accuracy, UA)、生产者精度(Producer Accuracy, PA)、总体精度(Overall Accuracy, OA)和 Kappa 系数。

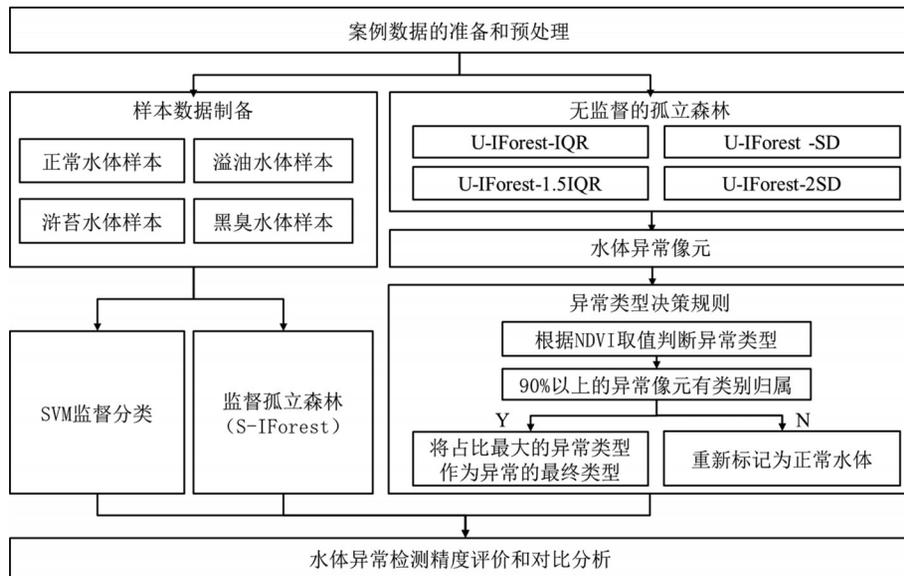


图3 水体异常发现和识别的流程

Fig.3 The basic flow of water anomaly discovery and identification

## 4 结果分析

图4展示了不同方法得到的3类异常水体检测的结果,表3为不同方法的检测精度统计。从表3可以看出本文所提出的方法(U-IForest-SD)相比SVM精度要低,但优于其他方法。U-IForest-SD方法对浒苔、黑臭水和溢油的总体识别精度分别为0.90, 0.96和0.94, Kappa系数分别为0.80, 0.90和0.85。阈值选择的方法对最终精度的影响很大,标准差阈值确定方法精度整体高于四分位距阈值确定方法的精度。标准差和四分位距中n的选择对结果影响大,n过大会导致漏分误差的增加,相应生产者精度会降低,这在黑臭水体和溢油的识别里面体现的尤为明显。当使用1.5倍IQR时,黑臭水的生产者精度只有0.04,使用2倍的标准差时黑臭水的生产者精度只有0.08,从而导致了其极低的Kappa系数。对比U-IForest-SD和UPCA-U-IForest-SD的结果可以看出,两者在浒苔的识别上基本没有差别,总体精度均为0.9, Kappa系数均为0.80;在黑臭水体上UPCA-U-IForest-SD的总体精度和Kappa系数略高于U-IForest-SD,而在溢油识别中U-IForest-SD的总体精度相比UPCA-U-IForest-SD高出了6%, Kappa系数高出了0.15。此外,在案例研究中使用原始哨兵数据12个波段进行异常检测比使用波段压缩后进行异常监测的耗时高出了约5倍。综合来看,处理前进行波段压缩是更好的选择。

图5显示了只用决策规则得到的3类异常水体

检测的结果,由图可以看出如果只使用决策规则,会出现大量的混分,即假警的情况,同一片水域会识别出多种不同的异常类型。而图6显示了基于正常水体的影像,仅使用U-IForest识别的异常水体。其中6.98%的正常水体被错误地识别为异常水体,但结合异常类型的决策规则后,所有的异常水体都被纠正为正常水体,避免了“虚警”的发生。由此可见,本文提出的U-IForest+决策规则的方法可以有效地避免“假警”和“虚警”发生。

## 5 讨论

### 5.1 本文所提方法的优势

本文所提的方法为基于单时相数据的非监督+决策规则的方法,该方法具有如下优点:

首先,整个算法只输入了单时相的数据,而传统的基于变化检测来发现和识别异常的方法往往要求多时相的数据,对数据质量的要求更高、处理步骤也更多更繁琐,致使在一定程度上降低了服务的即时性。

其次,非监督无需训练样本,提高了算法的可移植性。监督分类需要建立样本集,而样本集又往往基于特定区域的特定案例和特定数据集建立,这样的样本数据很难在不同区域及不同的遥感数据中复用,例如基于Landsat 8的7个多光谱波段数据建立的样本集,并不能很好的适用于有12个波段的Sentinel-2数据,而若引入大量不同的卫星数据制作样本需要投入大量人力、物力、时间和经费。

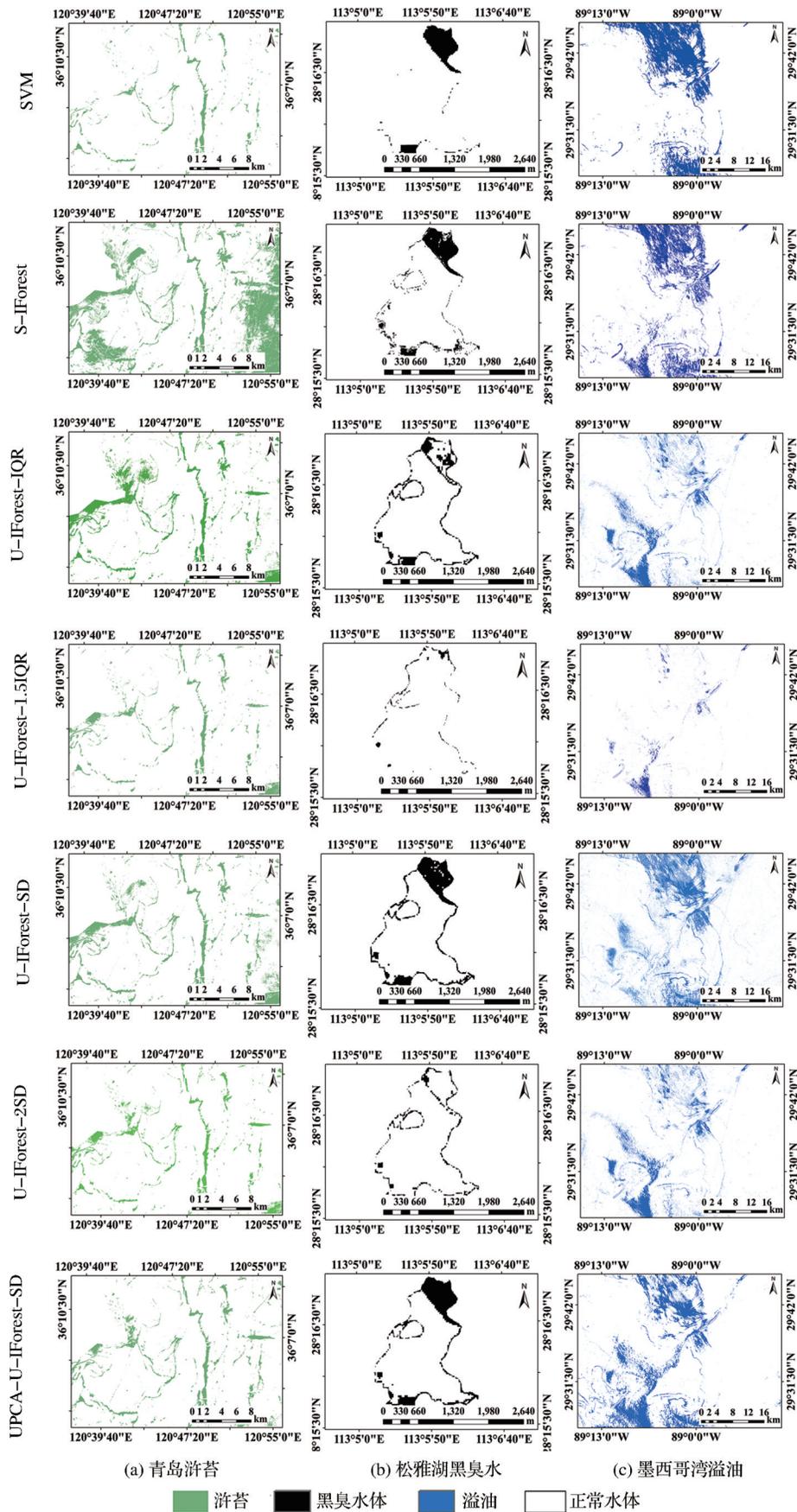


图 4 基于SVM、S-IForest和U-IForest 3类方法的异常水体识别结果

Fig.4 Abnormal water body identification results based on SVM, S-IForest and U-IForest methods

表3 精度验证

Table 3 Accuracy verification

案例	方法	PA	UA	OA	Kappa
青岛浒苔	SVM	0.94	0.99	0.97	0.95
	S-IForest	0.84	0.90	0.90	0.79
	U-IForest- IQR	0.76	0.95	0.88	0.75
	U-IForest- 1.5IQR	0.76	0.96	0.89	0.77
	U-IForest- SD	0.78	0.99	0.90	0.80
	U-IForest- 2SD	0.72	0.96	0.87	0.72
	UPCA-U-IForest-SD	0.80	0.96	0.90	0.80
松雅湖 黑臭水	SVM	0.99	0.99	0.99	0.99
	S-IForest	0.99	0.85	0.94	0.87
	U-IForest- IQR	0.43	0.86	0.80	0.46
	U-IForest- 1.5IQR	0.04	0.45	0.67	0.02
	U-IForest- SD	0.97	0.97	0.96	0.90
	U-IForest- 2SD	0.08	0.55	0.68	0.07
	UPCA-U-IForest-SD	0.99	0.93	0.97	0.95
墨西哥湾 溢油	SVM	0.91	0.99	0.98	0.94
	S-IForest	0.47	0.97	0.86	0.56
	U-IForest- IQR	0.34	0.86	0.77	0.38
	U-IForest- 1.5IQR	0.11	0.97	0.80	0.17
	U-IForest- SD	0.84	0.95	0.94	0.85
	U-IForest- 2SD	0.33	0.86	0.77	0.37
	UPCA-U-IForest-SD	0.80	0.76	0.88	0.70

再次,在制定决策规则时,考虑到不同传感器波段设置的差异,即便对于同类型的波段(如红波段),其具体的波段的波长区间范围也不完全一致,但大多数传感器都具有近红外和红光波段,换算成NDVI后可以增强不同传感器间的可比性,使得决策规则更具有普适性和稳定性。

最后,在实时水体异常监测时,对于是否有异常发生(异常发生的位置)以及异常的类型(浒苔、黑臭水和溢油)是不知道,但基于以往的经验,在特定的水域,其可能发生的水体异常事件的类型在一

定情况下是可以预判的。为此,会存在两个业务情景:①对于有先验知识的特定区域,例如青岛浒苔近年来时有发生,在该特定水域可以进行指定类型(浒苔)的异常水体监测;②对于没有先验知识的区域,需要同时进行各种异常水体类型(本文中特指浒苔、黑臭水和溢油)的监测。无论是哪种情景该算法都可以适用,且整个算法无需人工干预,在确定监测区域后,只要获取到监测区的图像,就可以按照本文设计的流程进行处理,提高了自动化程度。

## 5.2 影响本文所提方法的精度的因素

研究所提出的方法主要涉及3个环节:水体掩模的制作、异常水体的发现和异常水体类型的判断。因此,影响各个环节精度的因素都会影响最终水体异常发现和诊断的结果精度。

首先,异常指的是样本中的一些数值明显偏离其余数值的样本点。水体异常则是指异常水体像元特征值(如反射率)明显偏离其余水体(正常水体)像元特征值的像元。水体掩模首先明确了水体异常检测的范围,使得检测任务有的放矢。如果没有水体掩模,在整幅图像上进行检测,相对于正常水体来说是异常的水体像元在整幅图像上可能并不属于异常点。以浒苔为例,其光谱反射信息接近于正常的植被,在不考虑背景信息的情况下,对单时相的整幅图像进行分类时,可能会分为植被,并无异常。但如果增加水体掩模,“植被信息”出现在水体中则可以确定有异常发生。因此,水体掩模的应用是必要的。水体掩模可以在开展实时监测任务之前提前制备,例如来自各种已有的高精度土地利用覆盖图,然而水体掩模的精度也必然会影响到后续的检测精度,例如在水体掩模中如果有遗漏的水体则后续该区域都会排除在检测之外。

其次,本研究基于非监督孤立森林方法进行水

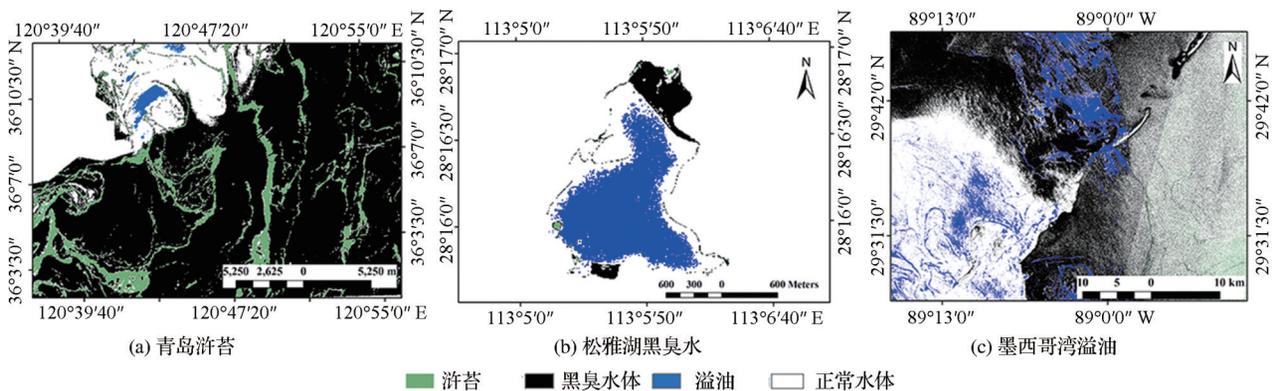


图5 仅依赖决策规则的异常水体识别结果

Fig.5 Abnormal water body identification results relying only on decision rules

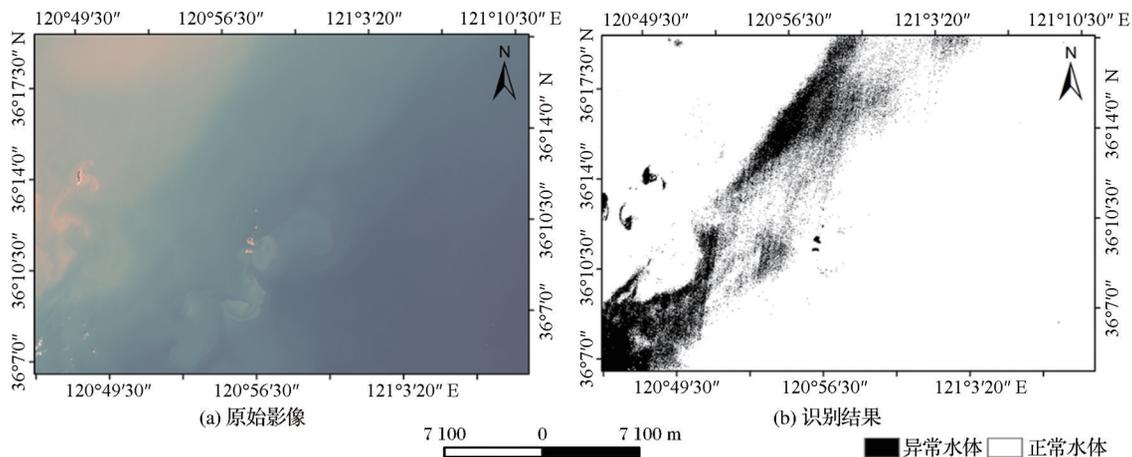


图6 仅使用U-IForest的异常水体识别结果

Fig.6 Abnormal water body identification results using U-IForest only

体异常的像元的快速发现。非监督孤立森林首先给出的是各个像元的异常程度,还需要通过一定阈值将其转化为异常非异常二值图。为了使算法具有更好的普适性,本文结合统计学对异常值的定义,选择了统计学异常值检测的两种常用方法(四分位距法和标准差法)来进行阈值的确定,经过测试标准差法精度更好。但标准差法中 $n$ 的选择,针对不同区域不同水体异常事件可能不同。在未来还需要进一步做测试。

最后,实验通过简单的决策规则来对水体异常类型进行判断。决策过程分为两步:①通过NDVI判断异常像元的异常类型;②对于一幅图像上的异常水体像元,统计各种异常水体类型像元的总数,将总数最大的异常水体类型作为该图像上异常水体的最终类型,将判读为其他异常类型的像元重编码至该类型。第一步中NDVI阈值的选择对结果有影响,目前我们只是在有限的数据集下进行的阈值设定和测试,未来需要收集更多的案例进行阈值适用性的评价和优化。第二步的一个潜在假设是对应一个监测区域,异常水体的类型只有一种,这在某些情况下可能并不适用。

## 6 结 论

本文提出了一种基于无监督的孤立森林(U-IForest)加决策规则的水体异常信息提取方法,并以Landsat与Sentinel的数据作为输入数据集,以青岛浒苔、松雅湖黑臭水、墨西哥湾溢油为案例,进行了算法的验证。研究结果显示:该方法对于3种异常类型的总体识别精度都在90%以上、Kappa系数都在0.8以上。尽管精度比SVM略低,但具有对数

据要求低、算法可移植性好、普适性强、自动化程度高的优点。此外,该方法可以有效地避免“假警”和“虚警”的发生,在水体异常的快速发现和识别业务中有很好的应用前景。

### 参考文献(References):

- [1] BLONDEAU-PATISSIER D, GOWER J F, DEKKER A G, *et al.* A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans[J]. *Progress in Oceanography*, 2014(123): 123-144.
- [2] WANG Xiaoqin, CHEN Chongcheng. Application of remote sensing to environmental monitoring in coastal waters[J]. *Marine Environmental Science*, 2000, 19(4): 72-76. [汪小钦, 陈崇成. 遥感在近岸海洋环境监测中的应用[J]. *海洋环境科学*, 2000, 19(4): 72-76.]
- [3] LIN Xiaochen, YANG Liangliang. Applications of satellite-based remote sensing techniques in water pollution monitoring [J]. *Sichuan Environment*, 2023, 42(1): 306-314 [林晓昇, 杨亮亮. 卫星影像遥感技术在水污染监测的应用[J]. *四川环境*, 2023, 42(1): 309-314.]
- [4] LOU Xiulin, HUANG Weigen. An artificial neural network method for detecting red tides with NOAA AVHRR imagery [J]. *Journal of Remote Sensing*, 2003, 8(2): 125-130, 162. [楼秀林, 黄韦良. 基于人工神经网络的赤潮卫星遥感方法研究[J]. *遥感学报*, 2003, 8(2): 125-130, 162.]
- [5] MICHELI F. Derivation of Red Tide index and density using Geostationary Ocean Color Imager (GOCI) data[J]. *Remote Sensing*, 2021, 13.D0I:10.3390/rs13020298
- [6] JIANG Dejuan, WANG Kun, XIA Yun. Comparative studies on remote sensing techniques for red tide monitoring in Bohai Sea [J]. *Marine Environmental Science*, 2020, 39(3): 460-467. [姜德娟, 王昆, 夏云. 渤海赤潮遥感监测方法比较研究[J]. *海洋环境科学*, 2020, 39(3): 460-467.]
- [7] HU C, MULLER-KARGER F E, TAYLOR C J, *et al.*

- Red tide detection and tracing using modis fluorescence data: A regional example in SW Florida coastal waters[J]. *Remote Sensing of Environment*, 2005, 97 (3) : 311-321. DOI: 10.1016/j.rse.2005.05.013
- [8] YE Na, JIA Jianjun, TIAN Jing, *et al.* Advances in the study of ulvapolifera monitoring with remote sensing [J]. *Remote Sensing for Land & Resources*, 2013, 25(1):7-12.[叶娜,贾建军,田静等.浒苔遥感监测方法的研究进展[J].*国土资源遥感*, 2013, 25(1):7-12.]
- [9] WANG Yiren, WANG Shengqiang, YU Yue, *et al.* An adaptive threshold algorithm for detecting ulva prokifera in Sounthern Yellow Sea by remote sensing[J].*Remote Sensing Information*, 2021, 36(2):120-129.[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.]
- [10] SCANLAN C M, FODEN J, WELLS E, *et al.* The monitoring of opportunistic macroalgal blooms for the water framework directive [J]. *Marine pollutionbulletin*, 2007, 55(1-6) : 162-167. DOI: 10.1016/j.marpolbul.2006.09.017
- [11] WU Shihong. Research progress of remote sensing monitoring key technologies for urban black and odorous water bodies[J]. *Chinese Journal of Environmental Engineering*, 2019, 13(6) : 1261-1271.[吴世红.城市黑臭水体遥感监测关键技术研究进展[J].*环境工程学报*, 2019, 13(6):1261-1271.]
- [12] CAO Yun, HANG Xin, GAO Yi, *et al.* Remote sensing monitoring of urban black and odorous water bodies using GF-2 images: Taking the main urban area of Nanjing as an example [J]. *Sichuan Environment*, 2023, 42(1) : 208-217.[曹云,杭鑫,高艺等.利用高分二号影像对城市黑臭水体遥感监测——以南京市主城区为例[J].*四川环境*, 2023, 42(1):208-217.]
- [13] LU Yingcheng, LIU Jianqiang, DING Jing, *et al.* Optical remote identification of spilled oils from the SANCHI oil tanker collision in the East China Sea [J]. *Chinese Science Bulletin*, 2019, 64(31) : 3213-3222.[陆应诚,刘建强,丁静等.中国东海“桑吉”轮溢油污染类型的光学遥感识别[J].*科学通报*, 2019, 64(31):3213-3222.]
- [14] REN Guangbo, GUO Jie, MA Yi, *et al.* Oil spill detection and slick thickness measurement Via UAV hyperspectral imaging[J]. *Acta Oceanologica Sinica*, 2019, 41(5) : 146-158.[任广波,过杰,马毅,等.海面溢油无人机高光谱遥感检测与厚度估算方法[J].*海洋学报*, 2019, 41(5):146-158.]
- [15] FINGAS M, BROWN C. Review of oil spill remote sensing [J]. *Marine Pollution Bulletin*, 2014, 83(1) : 9-23. DOI: 10.1016/S1353-2561(98)00023-1
- [16] SU H, WU Z, ZHANG H, *et al.* Hyperspectral anomaly detection: A survey [J]. *IEEE Geoscience and Remote Sensing Magazine*, 2021, 10(1) : 64-90. DOI: 10.1109/MGRS. 2021. 3105440
- [17] TAO X, ZHENG Y, CHEN W, *et al.* SVDD-based weighted oversampling technique for imbalanced and overlapped dataset learning [J]. *Information Sciences: An International Journal*, 2022(588):13-51.
- [18] WEI S A. Flexible region of interest extraction algorithm with adaptive threshold for 3-D synthetic aperture radar images [J]. *Remote Sensing*, 2021, 13 (21) : 4308. DOI: 10.3390/rs 13214308
- [19] PENG M. Adaptive subspace signal detection in structured interference plus compound Gaussian Sea clutter [J]. *Remote Sensing*, 2022, 14(9) : 2274. DOI:10.3390/rs14092274
- [20] YANG Yang, LI Tiekun, YANG Shuwen, *et al.* Change detection of GF-1 remote sensing image based on spatial fuzzy C-means clustering and Bayesian Network [J]. *Geomatics & Spatial Information Technology*, 2023, 46(4) : 34-37, 42[杨洋,李轶鲲,杨树文等.基于空间模糊C均值聚类 and 贝叶斯网络的高分一号遥感影像变化检测[J].*测绘与空间地理信息*, 2023, 46(4):34-37, 42.]
- [21] WANG Qiao. Research framework of remote sensing monitoring and real-time diagnosis of earth surface anomalies [J]. *Acta Geodaetica et Cartographica Sinica*, 2022, 51 (7) : 1141-1152.[王桥.地表异常遥感探测与即时诊断方法研究框架 [J].*测绘学报*, 2022, 51(7):1141-1152.]
- [22] LIU F T, TING K M, ZHOU Z H. Isolation-based anomaly detection [J]. *ACM Transactions on Knowledge Discovery from Data*, 2012, 6(1):1-39. DOI:10.1145/2133360.2133363
- [23] SONG Xiangyu. Research on anomaly detection in hyperspectral remote sensing images by Isolation Forest [D]. ChangChung: Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, 2022.[宋向宇.基于孤立森林算法的高光谱遥感图像异常目标检测方法研究 [D].长春:中国科学院大学(中国科学院长春光学精密机械与物理研究所), 2022.]
- [24] XUE Yuanyuan, HUANG Yuancheng, SU Yuanchao. Hyperspectral anomaly detection based on isolation forest with spatial weighting [J]. *Science of Surveying and Mapping*, 2021, 46(7) : 92-98.[薛园园,黄远程,苏远超.空间加权的孤立森林高光谱影像异常目标检测 [J].*测绘科学*, 2021, 46(7):92-98.]
- [25] LIU Lu, LUO Nianxue, ZHAO Qiansheng. Prediction of the outbreak scale of enteromorpha prolifera in the Yellow Sea based on historical data [J]. *Surveying and Mapping Bulletin*, 2022(7) : 7-11.[刘璐,罗年学,赵前胜.基于历史数据的黄海浒苔爆发规模预测 [J].*测绘通报*, 2022(7):7-11.]
- [26] CHEN Shuai. Remote sensing recognition of black and odorous water bodies based on Landsat 8 Images-A case study in Changsha [D]. ChangSha: Changsha University of Science and Technology, 2021.[陈帅.基于 Landsat 8 影像的黑臭水体遥感识别 [D].长沙:长沙理工大学, 2021.]
- [27] LIAO Guoxiang. Numerical simulation of the transport and diffusion of spilled oil released from ‘Deepwater Horizon’ accident in the gulf of mexico [J]. *Ocean Development and Management*, 2022, 39(4) : 89-96.[廖国祥.“深水地平线”事故深海溢油输移扩散的数值模拟 [J].*海洋开发与管理*, 2022, 39(4):89-96.]
- [28] The Ministry of Housing and Urban-Rural Development of

- the People's Republic of China. Notice of the General Office of the Ministry of Housing and Urban-Rural Development and the General Office of the Ministry of Environmental Protection on Announcing the Investigation of Black and Odorous Water Bodies in Cities Across the Country[EB/OL] [https://www.mohurd.gov.cn/gongkai/fdzdgknr/tzgg/201602/20160219\\_226677.html](https://www.mohurd.gov.cn/gongkai/fdzdgknr/tzgg/201602/20160219_226677.html), 2016, 2022. [中华人民共和国住房和城乡建设部. 住房和城乡建设部办公厅 环境保护部办公厅关于公布全国城市黑臭水体排查情况的通知[EB/OL] [https://www.mohurd.gov.cn/gongkai/fdzdgknr/tzgg/201602/20160219\\_226677.html](https://www.mohurd.gov.cn/gongkai/fdzdgknr/tzgg/201602/20160219_226677.html), 2016, 2022.]
- [29] JIA Shaolin. The United states oil foreign policy to latin america during World War II[D]. Kaifeng: Henan University, 2020. [贾少林. 二战时期美国对拉丁美洲的石油外交政策[D]. 开封:河南大学, 2020.]
- [30] WARD C H, TUNNELL J W. Habitats and biota of the Gulf of Mexico: An overview[R]. Habitats and Biota of the Gulf of Mexico: Before the Deepwater Horizon Oil Spill: Volume 1: Water Quality, Sediments, Sediment Contaminants, Oil and Gas Seeps, Coastal Habitats, Offshore Plankton and Benthos, and Shellfish, 2017: 1-54.
- [31] ZHAO Dong. multi-spectral remote sensing technologies slicks based on hyperspectral and researches on identifying sea surface oil [D]. Wuhan: China University of Geosciences, 2019. [赵冬. 基于高/多光谱遥感技术的海表油膜识别方法研究[D]. 武汉:中国地质大学, 2019.]
- [32] LU Y, LI X, TIAN Q, *et al.* Progress in marine oil spill optical remote sensing: Detected targets, spectral response characteristics, and theories[J]. Marine Geodesy, 2013, 36(3): 334-346.
- [33] LIU F T, TING K M, ZHOU Z H. Isolation Forest[C]// Proceedings of the 2008 IEEE International Conference on Data Mining. IEEE Computer Society, 2008.
- [34] CAO Y. A method based on improved IForest for trunk extraction and denoising of individual street trees [J]. Remote Sensing, 2022, 15(1): 115. DOI:10.3390/rs15010115

## Rapid Detection and Identification of Water Anomalies based on Isolated Forest

ZHU Xiufang<sup>1,2</sup>, LI Yuan<sup>3,4</sup>, GUO Rui<sup>1,2</sup>

(1.State Key Laboratory of Remote Sensing Science, Beijing Normal University, Beijing 100875, China;

2.Institute of Remote Sensing Science and Engineering, School of Geographic Sciences, Beijing Normal University, Beijing 100875, China;

3.College of Ecology and Environment, Inner Mongolia University, Hohhot 010021, China;

4.Key Laboratory of River and Lake Ecology, Inner Mongolia Autonomous Region, Hohhot 010021, China)

**Abstract:** The existing detection research of abnormal water bodies is usually carried out for specific regions, specific data sources and specific time phases. Anomaly recognition algorithm testing is often a backtracking of the water body anomaly events that have occurred, rather than real-time monitoring of the anomaly events, which cannot serve the requirements of rapid detection and identification of water body anomalies. In this paper, a method of extracting water body abnormal information based on unsupervised isolated forest plus decision rule (U-IForest-SD) is proposed. We selected Landsat and Sentinel as the test data, and tested the accuracy of U-IForest-SD with the black and smelly water body of Qingdao Enteromorpha, Songya lake and the oil spill in the Gulf of Mexico as research cases. We also compared U-IForest-SD with SVM and supervised isolated forests. The results show that the overall accuracy of the proposed method for the three types of anomalies is above 90%, and the kappa coefficient is above 0.8. The overall accuracy is higher than that of supervised isolated forest but slightly lower than that of SVM. This algorithm only needs to input single phase images, and does not need training samples. It has the advantages of good portability, strong universality and high automation. In addition, it can effectively avoid the occurrence of "wrong alarm" and "false alarm". Therefore, the newly proposed method has a good application prospect in the rapid detection and identification of abnormal water bodies.

**Key words:** Anomaly detection; Oil spill; Black smelly water; Enteromorpha