

大气校正算法对高光谱反演水体叶绿素 a 浓度的影响

孟祥亮^{1,2}, 冯建飞¹, 付萍杰^{1,*}, 张家威³, 张雨焯¹, 孟飞¹

(1. 山东建筑大学 测绘地理信息学院, 山东 济南 250101; 2. 山东省生态环境监测中心, 山东 济南 250101; 3. 山东科技大学 测绘与空间信息学院, 山东 青岛 266000)

摘要: 针对水体叶绿素 a 浓度监测时大气介质影响光谱真实性的问题, 文章从影像外部大气产品校正、内部大气补偿参数校正和光谱均值校正等方面分别对资源 1 号 02D (ZY-1 02D) 高光谱卫星影像进行大气校正以突出水体信号, 同时借助多维光谱指数和 CatBoost 机器学习算法进一步提高水体叶绿素 a 浓度的反演精度。结果表明: 大气校正算法在独山湖的应用中, 6S 优于 QUAC, 而 FLAASH 最差; CatBoost 模型能够更好地拟合预测误差, 提高反演精度; 6S 算法-四波段参数-CatBoost 模型的反演组合效果最好 ($R^2=0.80$)。

关键词: ZY-1 02D 高光谱影像; 大气校正; CatBoost; 南四湖; 叶绿素 a

中图分类号: P237

文献标识码: A

文章编号: 1673-7644(2024)01-0098-10

Influence of atmospheric correction algorithm on hyperspectral retrieval of chlorophyll-a in water

MENG Xiangliang^{1,2}, FENG Jianfei¹, FU Pingjie^{1,*}, ZHANG Jiawei³, ZHANG Yuxuan¹, MENG Fei¹

(1. School of Surveying, Mapping and Geographic Information, Shandong Jianzhu University, Jinan 250101, China; 2. Shandong Ecological Environment Monitoring Center, Jinan 250101, China; 3. School of Surveying and Spatial Information, Shandong University of Science and Technology, Qingdao 266000, China)

Abstract: In the process of monitoring the concentration of chlorophyll-a in water, the destruction of the spectral authenticity by the atmospheric medium is a difficult problem to be solved. In response to this problem, this study focuses on atmospheric correction of ZY-1 02D hyperspectral satellite imagery from three perspectives; external atmospheric products, internal atmospheric compensation parameters and spectral mean of the image, so as to enhance water signal accuracy; At the same time, with the help of multi-dimensional spectral index, combined with the CatBoost machine learning algorithm, the inversion accuracy of water chlorophyll-a concentration is further improved. The results show that in the application of atmospheric correction algorithm in Dushan Lake, 6S is better than QUAC, and FLAASH is the worst. CatBoost model can better fit the prediction error and improve the inversion accuracy. 6S algorithm-Four-band parameter-CatBoost model inversion combination works best (R^2 reaches 0.80).

Key words: ZY-1 02D hyperspectral image; atmospheric correction; CatBoost; Nansihu Lake; chlorophyll-a

收稿日期: 2022-08-31

基金项目: 国家自然科学基金项目(42101388)

作者简介: 孟祥亮(1981—), 男, 高级工程师, 硕士, 主要从事生态环境监测与评价等方面的研究。E-mail: 24692410@qq.com

通信作者*: 付萍杰(1989—), 女, 讲师, 博士, 主要从事高光谱遥感与环境监测等方面的研究。E-mail: fupjrs@126.com

0 引言

水体叶绿素 a (Chla) 是水体富营养化程度的重要指标之一,对湖库叶绿素 a 浓度的广域定量监测和实时定点分析是监测湖库水体营养状态的基本需求。水体叶绿素 a 浓度的监测难点之一在于其空间异质性,地面监测虽然精度较高,但局限于某一特定时刻,难以满足动态变化的监测。遥感技术具有监测范围广、监测时序长的优势^[1],但大气中水汽分子、气溶胶等物质会消耗掉大部分的离水辐射信号(90%~95%),且内陆水体空间布局分散,地理状况多变,其光学特性表现为复杂性和区域性^[2-4]。大气校正能够消除部分大气影响,还原目标地物的光谱信号,高精度的大气校正可以降低不同空间和不同时相影像大气校正误差对模型构建和应用带来的不确定性,是提升水质遥感监测模型时空移植性的前提,对于提高水色反演模型的准确性和普适性具有重要意义。

学者们针对水质反演的大气校正方法开展了多项研究。杜挺等^[5]采用 HJ-1B 多光谱影像,对太湖流域进行的反演中发现光谱超立方体的快速视线大气分析(Fast Line-of-Fight Atmospheric Analysis of Spectral Hypercubes, FLAASH)和对太阳光谱中卫星信号的二次分析(Second Simulation of the Satellite Signal in the Solar Spectrum, 6S)算法效果较好,但快速大气校正(Quick Atmospheric Correction, QUAC)算法下的典型地物光谱出现失真现象。商子健等^[6]以归一化植被指数(Normalized Difference Vegetation Index, NDVI)为指标,验证了 6S 大气校正模型在 GF-1 多光谱影像水体反演方面的适用性,其效果优于 FLAASH 大气校正。曾群等^[7]将不同的波段组合因子运用到大气校正模型中,发现对于浑浊、高动态特性的水体,FLAASH 优于 QUAC。潘岑岑等^[8]利用 Hypesx 高光谱数据的研究指出,基于统计学的 QUAC 和经验线性法两种大气校正方法效果不稳定,导致其系数变化较大。崔成岭等^[9]指出 6S 大气校正查找表精度较低的现状,通过实时创建查找表的方式对资源 1 号 02D(ZY-1 02D)高光

谱影像做大气校正处理,总体精度与 FLAASH 相近,且发现 6S 对蓝光波段的校正不完全。刘瑶等^[10]测试了资源 1 号 02D 高光谱影像在内陆水体叶绿素 a 浓度反演方面的适用性,其波段比值模型取得了最好的效果,并指出针对于 ZY-1 02D 水体图像的降噪和大气校正方法是未来的发展方向。高光谱影像的大气校正正在针对不同地物的反演中亦表现出不同的效果^[11-12],了解地物反演的参数贡献有利于提高大气校正的精度和效率。目前,针对高光谱影像在二类水体叶绿素 a 浓度反演方面的大气校正方法相关研究较少,由于波段连续细微的高光谱数据对大气干扰极为敏感,需要更精准的大气校正处理,才能满足水体叶绿素 a 浓度反演的数据需求。

文章结合内陆水环境研究对影像光谱高精度的需求及高光谱卫星遥感器的特点,以南四湖为研究区,以(ZY-1 02D)高光谱遥感影像为数据源,按照 6S、FLAASH、QUAC 大气校正的特点,从影像外部大气产品校正、影像内部大气补偿参数校正和影像的光谱均值校正 3 个角度分别对影像进行大气校正,进而提取多维光谱指数,并利用半分析和机器学习方法验证反演模型的反演精度。

1 数据源与实验区选取

1.1 研究区和水体样本采集

南四湖位于微山县境内,由微山、昭阳、独山和南阳等 4 个湖泊南北相连组成,是南水北调东线工程重要的水源地^[13-14],湖内水生生物与水禽种类众多^[15],其水质状况对东线工程以及生物多样性影响巨大。南四湖面积约为 1 266 km²,平均水深为 1.5 m^[16],其流域多年平均年降水量在 700 mm 以上,入湖河流有 50 多条^[17],且来水河流近九成集中于上级湖,外流入湖和湖内运作均对南四湖水质产生不同程度影响^[18]。

根据卫星过境时间,按照均匀布点原则,在独山湖周边采集 38 个表层水样,分布如图 1 所示。文章所用地图审图号为鲁 SG(2023)031 号。

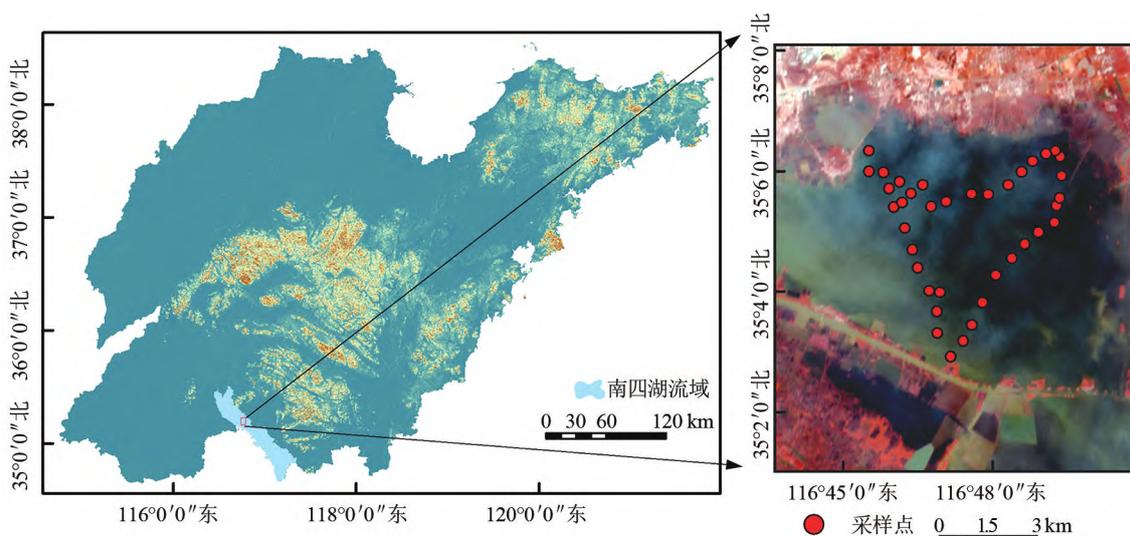


图1 南四湖位置及实测点分布示意图

1.2 数据源及其预处理

遥感影像数据为2021年9月12日的ZY-1 02D高光谱数据,其搭载的可见短波红外高光谱相机(Advanced Hyperspectral Imager, AHSI)传感器比高分5号的信噪比更高,但波段数量减少到166个,光谱分辨率也随之降低^[19]。原始影像数据经辐射定标处理,得到大气校正需要的表观反射率和辐射亮度值。

外业数据采集利用美国ASD公司生产的FieldSpec 4 Hi-Res便携式地物光谱仪,按照白板、水体、天空、白板的顺序,实测采样点水体高光谱数据,每点重复5次。瓶装湖面表层水样^[20],放入黑色不透光袋中保存,并用GPS定位仪记录采集点位置信息。返回室内避光处,采用美国安诺牌ChloroTech 121A手持式叶绿素测定仪测定水体叶绿素a浓度。

2 理论和方法

2.1 大气校正模型

大气校正可以保留由影像地物重要组成成分差异导致的反射率的微小差别信息^[8]。针对不同地物的光谱异质性及周边环境的差异,衍生出的大气校正方法也具有各自的侧重性。6S模型侧重影像过境时的观测及大气条件参数,并基于此参数建立的查找表进行逐像元插值得到监测点的大气校正系数;FLAASH模型则侧重于影像像素光谱特征,考虑周边像元对目标像元造成的“邻近效应”、光谱噪声等因素对目标地物校正;QUAC模型则是整幅影像的光谱信息收集,监测点的光谱特征即为目标地物

的光谱均值。

6S辐射传输模型由5S模型发展而来^[21],波段处理范围为0.25~4.0 μm,通过模拟信号传输过程的太阳辐射,并计算信号在进入传感器前的辐射能量,得到校正参数进行大气校正^[7]。ZY-1 02D高光谱影像需进行可见光近红外波段(Visible and Near-Infrared, VNIR)和短波红外波段(Short Wave Infrared, SWIR)的数据整合,将表观辐射亮度转换为表观反射率,由式(1)表示为

$$\rho_{\text{TOA}} = \frac{\pi \times L_{\lambda} \times d^2}{E_{\lambda} \times \cos\theta}, \quad (1)$$

式中 ρ_{TOA} 为表观反射率; L_{λ} 为波段 λ 处的中心波长, nm; d 为日地距离, AU; E_{λ} 为波段 λ 处的太阳光谱辐照度, $\text{W}/(\text{m}^2 \cdot \mu\text{m})$; θ 为太阳天顶角, ($^{\circ}$)。

6S模型通过输入表观反射率和影像元文件,得到各波段大气校正系数 x_a, x_b, x_c 。将表观反射率转化为真实的地表反射率 ρ_r ,可由式(2)和(3)表示为

$$\rho_r = \frac{y}{1 + x_c \times y}, \quad (2)$$

$$y = x_a \times L_{\lambda} - x_b. \quad (3)$$

FLAASH模型基于MODTRAN4+辐射传输模型,波段区间0.4~2.5 μm。通过设定模型参数,逐像元反演校正参数。针对不同的波段区间,进行特定的水汽反演,采用暗目标法进行气溶胶厚度的反演。影像像元光谱辐射亮度由式(4)表示为

$$L = \frac{A\rho}{1 - \rho_e S} + \frac{B\rho_e}{1 - \rho_e S} + L_{\alpha}, \quad (4)$$

式中 L 为像元辐射亮度; ρ 和 ρ_e 分别为像元与相邻像元地表反射率均值; S 为大气球面反照率; L_{α} 为大气程辐射; A, B 是依赖于大气(透过率)和几何状况的

系数。 S 、 L_α 、 A 、 B 可由辐射传输模型 MODTRAN4+ 计算得到。

QUAC 是基于影像本身的统计方法,采集像元内的地物光谱值,取同一地物的光谱平均值作为判别地物的经验值,利用端元平均反射率与参考物进行对比确定大气的影。QUAC 大气校正算法并不依赖影像获取过程中的各类参数信息,大气参数和仪器标定的误差对校正结果的影响较小^[7]。端元平均反射率 ρ' 由式(5)表示为

$$\rho' = (\rho_1 + \rho_2 + \dots + \rho_n) / n, \quad (5)$$

式中 $\rho_1, \rho_2, \dots, \rho_n$ 为视场内各物质的端元光谱反射率; n 为端元个数。

2.2 水体叶绿素 a 浓度反演

(1) 三波段光谱指数

光谱指数是一种基于光谱反射率构建的指标函数,依据地物的光谱特性,通过波段组合度量地表参量^[22]。在尝试了多种三波段组合后,选择遍历后相关系数较高的公式反演三波段指数(Three-Band Index, TBI),由式(6)表示为

$$TBI = (R_{\lambda_1} - R_{\lambda_2}) / (R_{\lambda_3} - R_{\lambda_2}), \quad (6)$$

式中 $R_{\lambda_1}, R_{\lambda_2}, R_{\lambda_3}$ 分别为 395~900 nm 范围内波长为 λ_1, λ_2 和 λ_3 处的遥感反射率。

(2) 四波段光谱指数

按照三波段光谱指数法选取波段组合的方法,运用水体反演叶绿素 a 浓度的四波段公式反演四波段指数(Four-Band Index, FBI),由式(7)^[23]表示为

$$FBI = (1/R_{\lambda_1} - 1/R_{\lambda_2}) / (1/R_{\lambda_4} - 1/R_{\lambda_3}), \quad (7)$$

式中 R_{λ_4} 为 395~900 nm 范围内波长为 λ_4 处的遥感反射率。

(3) CatBoost 算法

CatBoost 算法是梯度提升决策树(Gradient Boosting Decision Tree, GBDT)算法框架的改进算法,运用排序提升方法构建模型,在不同的迭代中会采用不同的排列顺序,训练集中的噪声点被消除,梯度估计与浓度预测误差得到了解决。CatBoost 算法通过减少超参数调优来抵抗模型过度拟合,增强了算法的准确性和泛化能力,其计算过程添加了先验值和权重参数,一些低频率类型的数据和噪声点对数据整体趋势的影响得到有效控制。CatBoost 算法可由式(8)表示为

$$x_{i,k} = \frac{\sum_{j=1}^{p-1} [x_{\sigma_j,k} = x_{\sigma_p,k}] Y_j + aP}{\sum_{j=1}^{p-1} [x_{\sigma_j,k} = x_{\sigma_p,k}] + a}, \quad (8)$$

式中 $x_{\sigma_j,k}, x_{\sigma_p,k}$ 分别为第 k 个训练样本的第 j, p 个类别特征; $x_{i,k}$ 为类别特征平均值; $[\]$ 为指示函数运算,即括号内两个量相等时取 1,否则取 0; Y_j 为第 j 个样本的标签; P 为添加的先验项; a 为先验值的权重。

(4) 反演模型构建

对大气校正的影像提取像元光谱,并基于叶绿素 a 实测浓度,提取最优 TBI 和 FBI 用于 CatBoost 模型的特征组建。CatBoost 模型采取相同的随机种子进行训练集(70%)与验证集(30%)的划分,网格搜索法在指定范围内进行枚举^[24],将效果最好的参数用于模型构建。

一元线性模型的建模参数分别为 TBI 和 FBI 的最优遍历结果^[25],训练集与验证集的划分同 CatBoost 模型。

2.3 大气校正及叶绿素 a 反演精度评价

为定量比较大气校正模型的校正效果,使用光谱角制图(Spectral Angle Mapper, SAM)^[26]、均方根误差(Root Mean Square Error, RMSE)和相关系数来衡量不同模型校正后的影像光谱反射率与实测反射率之间的接近程度^[21]。SAM 光谱角 α 的余弦值由式(9)表示为

$$\cos \alpha = \frac{A \cdot B}{|A| |B|} = \frac{\sum_{i=1}^N A_i B_i}{\sqrt{\sum_{i=1}^N A_i A_i} \sqrt{\sum_{i=1}^N B_i B_i}}, \quad (9)$$

式中 N 为样本总数; A_i, B_i 分别为第 i 个像元向量的光谱值。

相关系数 r 由式(10)表示为

$$r = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}}, \quad (10)$$

式中 X_i, Y_i 分别为像元光谱值和实测光谱值; \bar{X} 和 \bar{Y} 分别为像元、实测的光谱均值。

叶绿素 a 浓度反演精度采用决定系数 R^2 和均方根误差 RMSE 评价。RMSE 和 R^2 分别由式(11)和(12)表示为

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_{ai} - X_{ai})^2}, \quad (11)$$

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (12)$$

式中 X_{ai} 、 Y_{ai} 分别为叶绿素 a 浓度的实测值和反演值； y_i 、 \hat{y}_i 分别为实测值和模型反演值； \bar{y} 为 N 个 y_i 的均值。

3 结果与分析

3.1 大气校正

3.1.1 大气校正下的光谱曲线对比

由于实测点位分布在独山湖区域,因此仅针对

南四湖的独山湖区域展开分析。不同大气校正参数见表 1,观测日期为 2021 年 10 月 22 日,过境时间为世界标准时间 (Universal Time Coordinated, UTC) 03:19:42,中心经纬度为 35.28°N、116.80°E。其中,FLAASH、QUAC 大气校正参数均在 ENVI 中完成,由于没有相应的传感器类型,所以统一选择“unknown sensor”。

表 1 大气校正参数表

大气校正方法	传感器高度 /km	空间分辨率 /m	大气模式	气溶胶模式	能见度 /m	地物高度 /m	太阳方位角 /(°)	卫星方位角 /(°)
6S	778	30	Mid-Latitude Summer	Rural	40	34.5	167.44	173.76
FLAASH	778	30	Mid-Latitude Summer	Rural	40	34.5	*	*
QUAC	778	30	*	*	*	*	*	*

注:表中“*”表示参数设置保持默认或未使用该参数参与校正计算。

为保证光谱信息的可信度,任取一点(14 号点)的实测值和全部监测点光谱均值,分别与大气校正光谱对比,结果如图 2 所示。由于叶绿素和类胡萝卜素的吸收,在 440 和 490 nm 附近形成两个小反射谷;藻类色素的低吸收以及水中悬浮物的散射,使得光谱曲线在 570 nm 附近形成了大反射峰;4 种光谱曲线均在 675 和 700 nm 附近分别形成了大反射谷和大反射峰,这与前人的研究结果一致^[27]。对比图 2(a)和(b)发现,除反射率略有升高外,光谱均值与 14 号点光谱曲线的整体趋势基本一致。由于

受到水面天空光等多种信号的影响,大气校正曲线普遍高于实测光谱曲线。校正方式的不同使得光谱反射率的值高于实测值的程度也不一样,6S 借助大气校正产品逐一计算得到每波段每像素的地表反射率值,光谱反射率最接近实测值;而 FLAASH 是基于整张图像^[4],且重采样减少了光谱的波段数量,对于蓝光波段的不完全校正使得 FLAASH 曲线出现部分负值。QUAC 则是基于影像本身,且水体占比比较大,有利于进行水体的光谱校正。

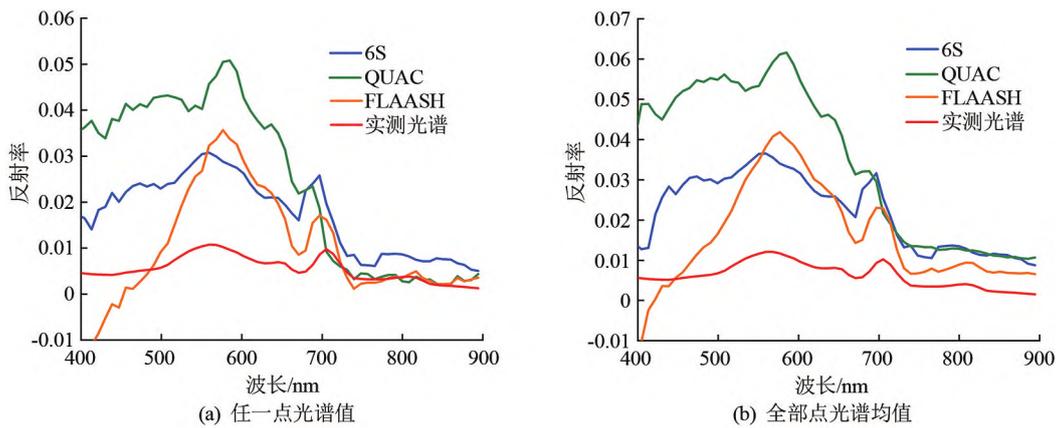


图 2 不同大气校正光谱对比

3.1.2 精度评价

38 个采样点的实测光谱数据与大气校正数据的指标值如图 3 所示。6S 和 QUAC 模型的光谱角余弦值都在 0.9 以上,而 FLAASH 模型的余弦值略低一些,且各点之间的差异较大。同时 6S 模型的 r 在 3 种大气校正模型中最高,这也说明 6S 模型对地

物光谱信息的保持度高。6S 与 FLAASH 模型的 RMSE 值均约为 0.15,QUAC 的值与实测值的离散程度较大。综上所述,6S 模型的评估结果均表现良好,FLAASH 模型的光谱信息保持较弱,QUAC 模型校正后的影像反射率缺乏代表性,校正效果不稳定。

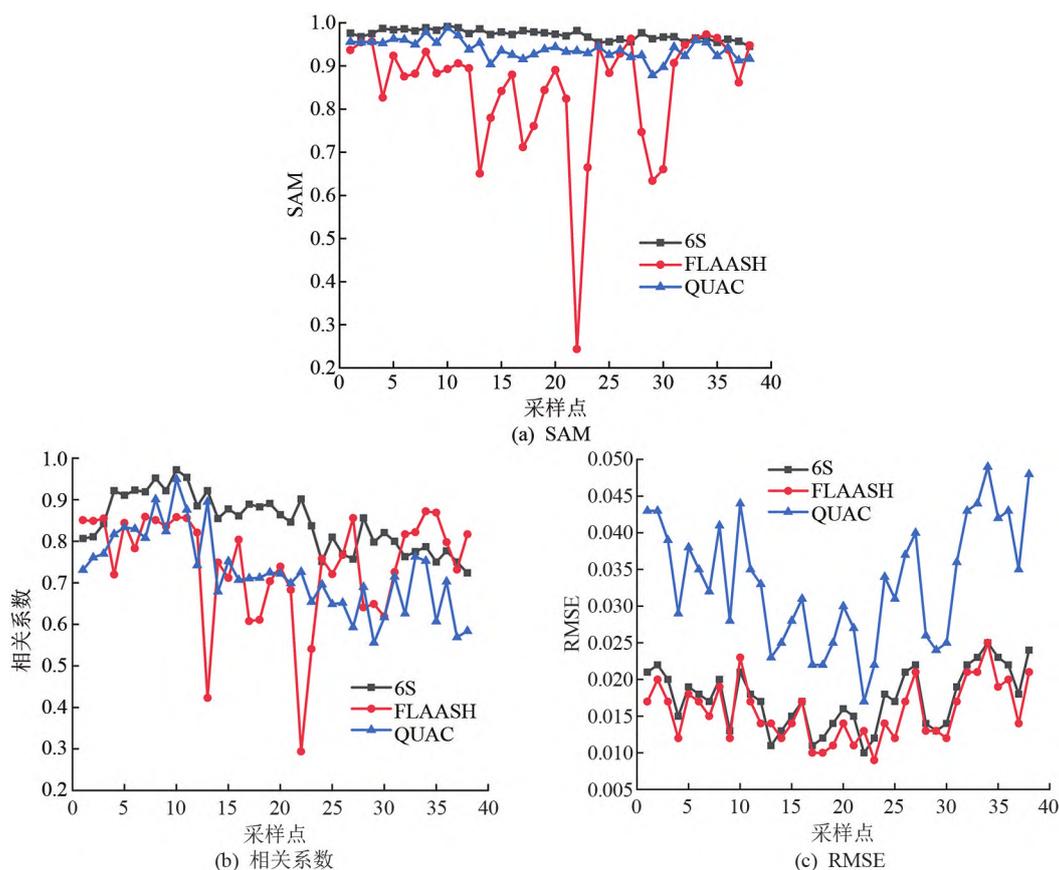


图 3 大气校正精度指标图

3.2 叶绿素 a 浓度反演

3.2.1 多维光谱指数提取

在文章研究的光谱波段范围内(395~900 nm),通过 MATLAB 软件实现基于 TBI 和 FBI 公式的波段反射率循环迭代。基于实测叶绿素 a 浓度,利用

相关系数 r 提取最优的波段组合,计算模型参数。选取相关性最高的前 5 个光谱指数组合作为 CatBoost 反演模型的参数,选取相关性最优的光谱指数进行一元线性回归模型的反演。TBI 和 FBI 的相关结果分别见表 2 和 3。

表 2 三波段光谱指数及相关系数

序号	6S 大气校正				FLAASH 大气校正				QUAC 大气校正			
	λ_1	λ_2	λ_3	r	λ_1	λ_2	λ_3	r	λ_1	λ_2	λ_3	r
1	696	542	395	0.644	825	722	524	0.701	404	567	473	0.621
2	696	542	404	0.636	825	722	533	0.685	473	404	559	0.562
3	542	395	696	0.630	722	524	825	0.683	490	404	559	0.556
4	542	404	696	0.619	722	533	825	0.670	816	859	808	0.550
5	395	696	542	0.607	876	722	524	0.652	413	567	473	0.547

表 3 四波段光谱指数及相关系数

序号	6S 大气校正					FLAASH 大气校正					QUAC 大气校正				
	λ_1	λ_2	λ_3	λ_4	r	λ_1	λ_2	λ_3	λ_4	r	λ_1	λ_2	λ_3	λ_4	r
1	507	499	645	730	0.673	876	576	765	516	0.637	576	533	696	799	0.682
2	507	499	636	730	0.673	876	567	765	516	0.634	533	576	808	696	0.670
3	507	499	628	730	0.670	876	559	765	516	0.634	576	533	645	833	0.664
4	507	499	619	730	0.667	876	593	765	516	0.632	576	533	628	833	0.662
5	507	499	653	730	0.665	765	516	576	876	0.632	576	533	696	833	0.659

3.2.2 叶绿素 a 浓度反演

对影像光谱数据和实测叶绿素 a 浓度,数据测

试集和训练集分别占 30%、70%。将基于 TBI 和 FBI 构建的一元线性反演模型应用到遥感影像上,

分别得到12个点的反演值。将不同算法的反演值与实测值进行拟合,结果如图4所示。6S大气校正后得到的叶绿素a浓度反演值,在以三波段和四波段作为特征参数的线性模型中,均具有良好表现。

QUAC大气校正与FLAASH大气校正后的线性模型反演精度略有差异,其中QUAC四波段线性模型精度最高。整体来说,反演值与实测值的 R^2 最高能达到0.74,模型效果较好。

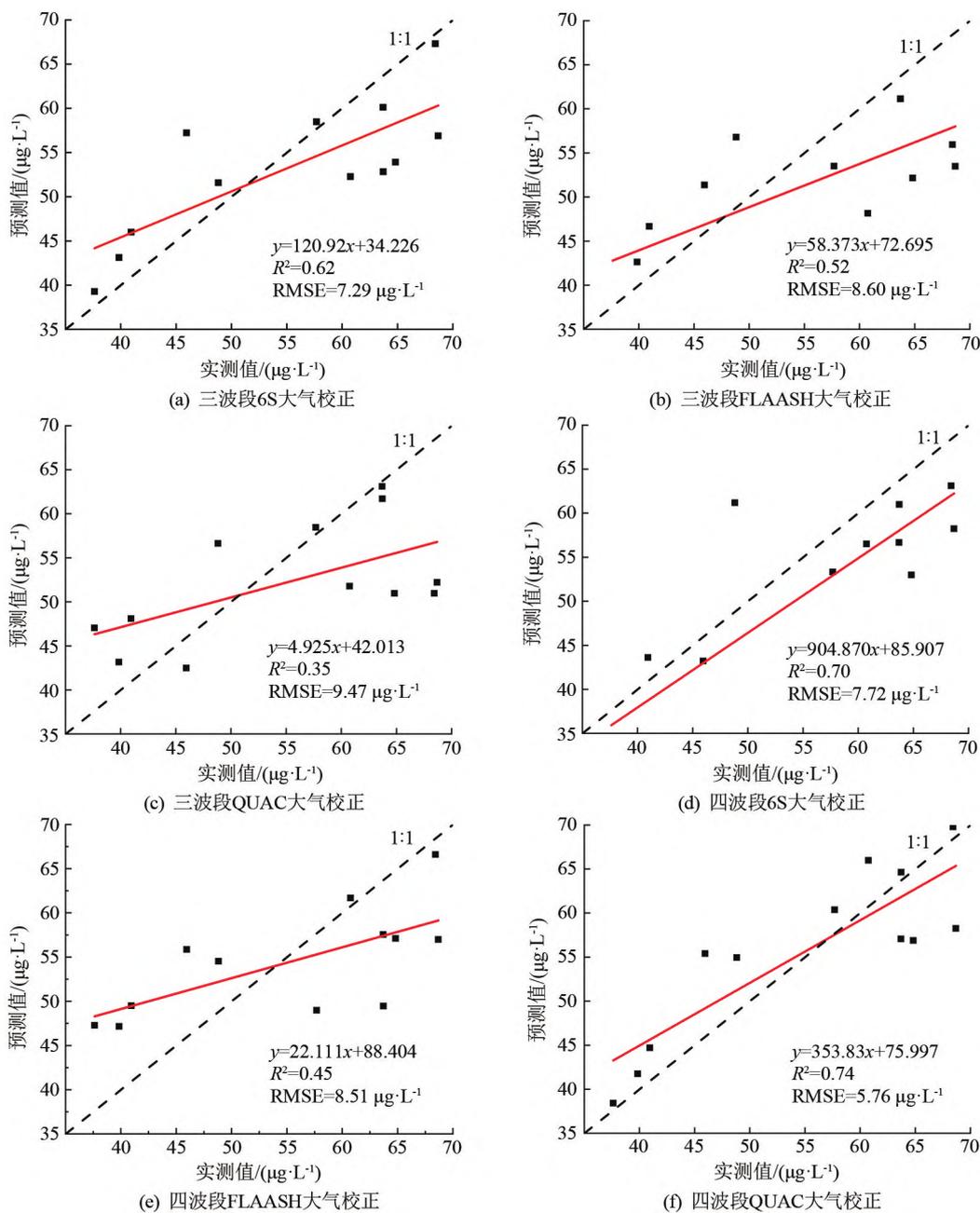


图4 不同大气校正算法反演结果与实测结果拟合图

选择与一元线性模型相同的训练集,由于CatBoost不需要过多调参,因此主要调节模型决策树的数量($n_estimators$)、学习率($learning_rate$)和决策树的深度($depth$),其余参数保持默认,按照参数调优标准,得到CatBoost模型参数见表4。

表4 CatBoost模型反演参数设置

算法	$n_estimators$		$learning_rate$		$depth$	
	三波段	四波段	三波段	四波段	三波段	四波段
6S	9	19	0.8	0.5	6	3
FLAASH	3	7	0.7	0.7	6	1
QUAC	3	12	0.9	0.5	5	2

AHSI 影像反演模型和精度评价结果见表 5。CatBoost 模型的建模精度总体良好,测试集最高相关系数 $R^2 = 0.80$ 、RMSE = 4.97 $\mu\text{g}/\text{L}$ 。由于 CatBoost 模型在不同参数组合中取得的精度最高,且在 6S 和 QUAC 大气校正后的参数反演结果均较好,因此选用 CatBoost 模型进行南四湖水体叶绿素 a 浓度反演。

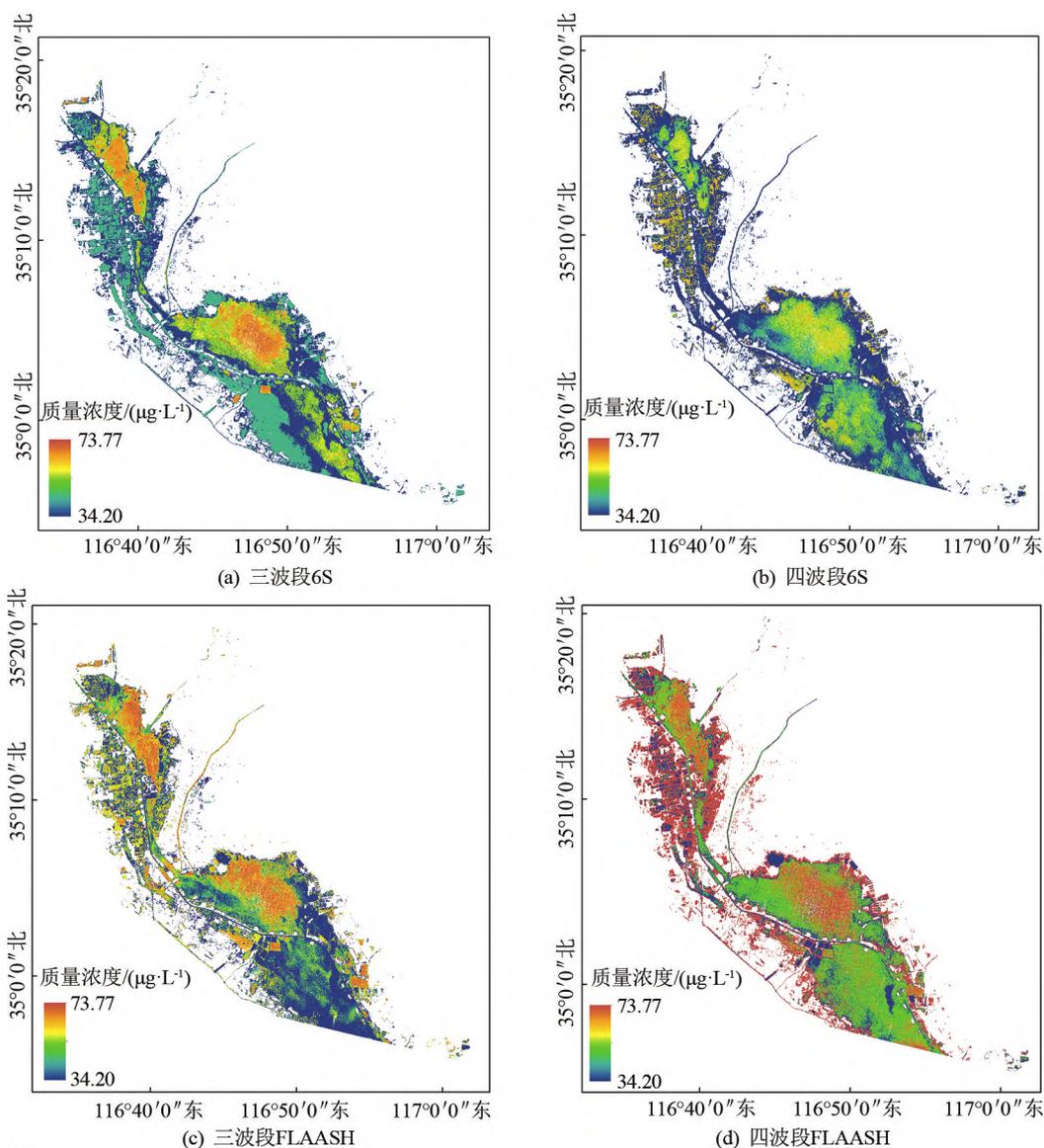
表 5 AHSI 影像反演模型和精度评价结果

算法	波段组合 方法	反演模型	R^2	RMSE /($\mu\text{g} \cdot \text{L}^{-1}$)
6S	三波段法	$y = 120.92x + 34.226$	0.62	7.29
		CatBoost	0.70	6.11
	四波段法	$y = 904.87x + 85.907$	0.70	7.72
		CatBoost	0.80	4.97
FLAASH	三波段法	$y = 58.373x + 72.695$	0.52	8.60
		CatBoost	0.37	8.92
	四波段法	$y = 22.111x + 88.404$	0.45	8.51
		CatBoost	0.37	8.89
QUAC	三波段法	$y = 4.925x + 42.013$	0.35	9.47
		CatBoost	0.63	6.78
	四波段法	$y = 353.83x + 75.997$	0.74	5.76
		CatBoost	0.76	5.46

对比不同大气校正的模型建模结果,6S 大气校正的模型拟合效果最好,其中四波段参数 CatBoost 模型的拟合精度最高。基于 QUAC 大气校正的三波段参数和四波段参数组合的 CatBoost 模型拟合精度均优于 FLAASH 大气校正。QUAC 大气校正效果缺乏稳定性,模型最低拟合精度 $R^2 = 0.35$ 。

将模型应用到影像,得到南四湖叶绿素 a 浓度反演结果如图 5 所示。可以看出,质量浓度空间差异明显,高值主要分布在南阳湖的东部沿岸和独山湖的湖心,流动性较弱区域,这与前人的研究结果相符^[28]。然而,该组合方法下模型的泛化能力较差,选取的波段对该地区的水体叶绿素 a 缺乏敏感性,影像模糊是由于反射信号包含高光谱噪声。

对比不同光谱指数的反演,四波段参数选择的波段反射率更接近实测反射率,参数信息更能体现叶绿素 a 浓度变化。这一点在 6S 与 QUAC 大气校正的影像中表现明显。



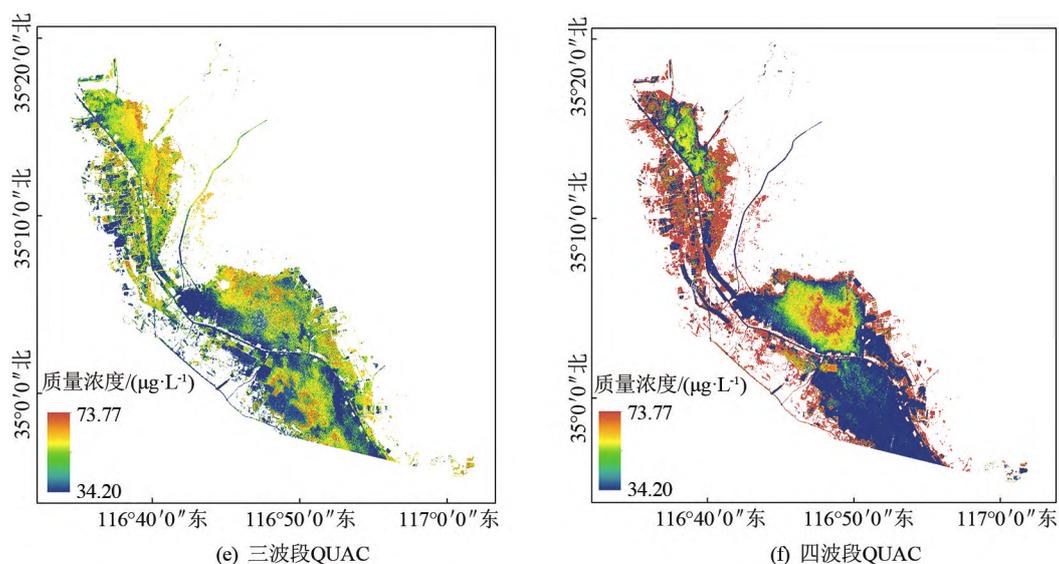


图5 不同大气校正下的 CatBoost 模型 AHSI 影像南四湖叶绿素 a 浓度

6S 大气校正的三波段与四波段参数模型反演, R^2 均稳定在较高水平, 表明 6S 大气校正对于高光谱反演水体叶绿素 a 浓度具有良好的适用性, 可作为影像预处理时的优先选择大气校正方法。

考虑到 QUAC 模型以影像本身作为对象, 对同类地物光谱采集取均值的校正特性, 对比叶绿素 a 浓度反演结果分析, 特征影像受地物异物同谱特性的影响较大。由于受大气等光学条件的影响, 不同水体叶绿素 a 浓度监测点的光谱可能存在较大差异^[5], 在求取均值后, 不同叶绿素 a 浓度对应的光谱值存在相近或相等的情况, 这将对反演造成困难, 三波段和四波段参数的光谱值出现误差, 造成最终反演精度出现不确定性。

FLAASH 大气校正后的反演效果均不理想, 可能是选取的 TBI 与 FBI 所用波段对叶绿素 a 并不敏感^[29], FLAASH 模型以大气校正产品作为系数, 不同波段参数的组合对影像的泛化能力也不尽相同, 导致构建的模型不能很好地预测水体中的叶绿素 a 浓度。

4 结论

基于南四湖 ZY-1 02D 影像, 分别进行了 6S、FLAASH 和 QUAC 的大气校正处理, 并对提取的光谱反射率进行多种组合, 分别利用线性回归模型和 CatBoost 模型中反演水体中叶绿素 a 浓度, 得到主要结论如下:

(1) 6S 模型校正效果最佳, QUAC 次之, FLAASH 模型效果最差;

(2) CatBoost 模型能在一定程度上提高反演精度, 排序提升的模型构建以及先验值的添加消除了误差噪声, 在不需要过多调参的情况下即可达到较好的回归预测精度;

(3) 6S 模型的四波段组合 CatBoost 模型反演结果的 R^2 最高为 0.80, 更有利于提高光谱数据在南四湖中部的独山湖叶绿素 a 浓度反演中的真实性和反演精度。

参考文献:

- [1] 陶然, 彭金婵, 张豪, 等. 内陆水体叶绿素 a 浓度遥感监测方法研究进展[J]. 地理信息世界, 2019, 26(4): 44-53.
- [2] 卢昭羿. 内陆水体多源卫星遥感大气校正方法研究[D]. 北京: 中国科学院大学(中国科学院遥感与数字地球研究所), 2019.
- [3] 张棋斐, 吴志峰, 解学通. 河口及近岸海域水体叶绿素浓度反演方法综述[J]. 生态科学, 2017, 36(5): 215-222.
- [4] 范文义, 孙晓芳, 王岩, 等. 基于两种辐射传输模型的遥感数据大气校正及结果对比分析[J]. 东北林业大学学报, 2009, 37(7): 121-124.
- [5] 杜挺, 王旭红, 梁海艳, 等. 湖泊叶绿素 a 反演的大气校正模型比较研究[J]. 云南大学学报(自然科学版), 2014, 36(2): 241-248.
- [6] 商子健, 金建文, 许振峰, 等. 高分一号多光谱影像 6S 大气校正研究[J]. 无线电工程, 2021, 51(8): 705-710.
- [7] 曾群, 张海东, 陈晓玲, 等. MERIS 影像水环境遥感大气校正算法评价——以鄱阳湖叶绿素 a 浓度反演为例[J]. 湖泊科学, 2016, 28(6): 1306-1315.

- [8] 潘岑岑,闫庆武,丁建伟,等.基于果蝇-鲍威尔优化的航空高光谱影像大气校正方法[J].光谱学与光谱分析,2018,38(1):224-234.
- [9] 崔成玲,张宏伟,范松滔,等.资源一号 02D 高光谱数据大气校正及应用[J].遥感信息,2022,37(5):45-50.
- [10] 刘瑶,李俊生,肖晨超,等.资源一号 02D 高光谱影像内陆水体叶绿素 a 浓度反演[J].遥感学报,2022,26(1):168-178.
- [11] 孔卓,杨海涛,郑逢杰,等.高光谱遥感图像大气校正研究进展[J].自然资源遥感,2022,34(4):1-10.
- [12] ZHANG J W, FU P J, MENG F, et al. Estimation algorithm for chlorophyll-a concentrations in water from hyperspectral images based on feature derivation and ensemble learning[J]. Ecological Informatics, 2022, 71: 101783.
- [13] 周冉,张志斌,张晓蕊,等.干湿交替对南四湖底泥真菌群落影响研究[J].山东建筑大学学报,2022,37(5):65-69.
- [14] 李利军,张国民,姚辉瑞,等.南四湖斜拉桥压重的优化研究[J].山东建筑大学学报,2022,37(4):121-127.
- [15] 于泉洲,梁春玲,张祖陆.近 40 年来南四湖湿地 NDVI 变化特征及其控制因子分析[J].湖泊科学,2014,26(3):455-463.
- [16] 杨安南,于泉洲,汤庆新,等.1984~2019 年南四湖水域变化特征及原因分析[J].水电能源科学,2021,39(7):49-52.
- [17] 张翠翠.南四湖对周边区域气候的影响[J].现代农业科技,2021(24):157-158.
- [18] 马景,武周虎.南四湖水质评价及改善效果分析[J].青岛理工大学学报,2021,42(6):101-105.
- [19] 刘银年,孙德新,韩波,等.资源一号 02D 卫星可见短波红外高光谱相机研制[J].航天器工程,2020,29(6):85-92.
- [20] 唐军武,田国良,汪小勇,等.水体光谱测量与分析 I:水面以上测量法[J].遥感学报,2004(1):37-44.
- [21] 刘梓钦,赵世湖,裴亮,等. GF-5 卫星高光谱数据大气校正反射率精度评价[J].遥感信息,2021,36(3):93-98.
- [22] 张子鹏,丁建丽,王敬哲,等.利用三维光谱指数定量估算土壤有机质含量:以新疆艾比湖流域为例[J].光谱学与光谱分析,2020,40(5):1514-1522.
- [23] LE C, LI Y, ZHA Y, et al. A four-band semi-analytical model for estimating chlorophyll a in highly turbid lakes: The case of Taihu Lake, China [J]. Remote Sensing of Environment, 2009, 113(6):1175-1182.
- [24] BERGSTRA J, BENGIO Y. Random search for hyper-parameter optimization[J]. Journal of Machine Learning Research, 2012, 13(1):281-305.
- [25] FENG L, HU C, HAN X, et al. Long-term distribution patterns of chlorophyll-a concentration in China's largest freshwater lake: MERIS full-resolution observations with a practical approach [J]. Remote Sensing, 2014, 7(1): 275-299.
- [26] 杜培军,陈云浩,方涛,等.高光谱遥感数据光谱特征的提取与应用[J].中国矿业大学学报,2003,32(5):500-504.
- [27] LIU G, LI L, SONG K, et al. An OLCI-based algorithm for semi-empirically partitioning absorption coefficient and estimating chlorophyll a concentration in various turbid case-2 waters [J]. Remote Sensing of Environment, 2020, 239:111648.
- [28] CUI Y, MENG F, FU P J, et al. Application of hyperspectral analysis of chlorophyll a concentration inversion in Nansi Lake [J]. Ecological Informatics, 2021, 64: 101360.
- [29] 潘应阳,国巧真,孙金华.水体叶绿素 a 浓度遥感反演方法研究进展[J].测绘科学,2017,42(1):43-48.

(学科责编:朱志鹏)